738

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

EDITED BY

CARL MURCHISON

Gordon Allport
Harvard University

Arthur G. Bills
University of Cincinnati

E. A. Bott
University of Toronto

Leonard Carmichael
Tufts College

Herbert S. Conrad
U. S. Office of Education

Elmer K. Culler
University of Rochester

John F. Dashiell
University of North Carolina

Roy M. Dorcus
University of California at Los Angeles

Horace B. English
Ohio State University

Franklin Fearing
University of California at Los Angeles

Frank A. Geldard
University of Virginia

Clarence H. Graham
Columbia University

Ernest R. Hilgard
Stanford University

H. M. Johnson
Tulane University

Heinrich Klüver
University of Chicago

Carney Landis
Columbia University

K. S. Lashley
Harvard University

Norman L. Munn
Bowdoin College

John Paul Nafe
Washington University

Donald G. Paterson
University of Minnesota

Lee Edward Travis
University of Southern California

Robert C. Tryon
University of California

Michael J. Zigler
Wellesley College

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

LIBRARY

VOLUME 35

1953

Copyright, 1953, by he Journal Press

Published Quarterly by the Journal Press
Provincetown, Massachusetts, U. S. A.

AUTHOR INDEX

Auble, Donavon 307
Bartley, S. Howard 299
Battersby, W. S. 329
Bender, M. B. 329
Bennett, Lillian 59
Block, Jack 50
Caldwell, Willard E. 353
Corbett, Gerald R. 81
Crannell, C. W. 131
Crosby, D. W. 229
Farber, Maurice L. 253
Fry, Franklyn D. 181, 197
Gilberg, R. 229
Grings, William W. 271
Grossack, Martin 241, 245
Guy, M. 201
Guze, Henry 313
Humphrey, Tryphena 3
Ittelson, William H. 235
Ives, Kenneth 45
Jarrett, R. F. 161
Kline, Milton V. 227
Knehr, G. A. 201
Lindgren, H. C. 229
McCarthy, Dorothea 155
Mech, E. Victor 283, 307
Newbury, Edward 371
Norris, Eugenia B. 107, 119
O'Brien, Cyril C. 267
Obrist, Walter D. 259
Perlmutter, Howard V. 361
Porteus, Stanley D. 81
Ruesch, Jurgen 59
Sherriffs, A. C. 161
Smith, William M. 143
Spragg, S. D. S. 107, 119
Teuber, H. L. 329
Tilton, J. W. 169
Vickery, A. 201
Wilkinson, Frank Ray 299
Womack, H. Lynn 353

Bur[illegible] Edn[illegible] Research
[illegible]ING COLLEGE
11.8.70
Dated
No J 114

TABLE OF CONTENTS

The relation of oxygen deprivation to fetal reflex arcs and the development of fetal behavior 3
By Tryphena Humphrey

The preparation of case balance sheets 45
By Kenneth Ives

The assessment of communication: I. A method for the analysis of social interaction 59
By Jurgen Ruesch, Jack Block, and Lillian Bennett

Statutory definitions of feebleminded in U. S. A. 81
By Stanley D. Porteus and Gerald R. Corbett

Performance on a following tracking task (modified *SAM* two-hand co-ordination test) as a function of the planes of operation of the controls 107
By Eugenia B. Norris and S. D. S. Spragg

Performance on a following tracking task (modified *SAM* two-hand co-ordination test) as a function of the relations between direction of rotation of controls and direction of movement of display 119
By Eugenia B Norris and S. D. S. Spragg

The validity of certain measures of art appreciation in relation to a drawing task 131
By C. W. Crannell

A methodological study of size-distance perception 143
By William M. Smith

Some possible explanations of sex differences in language development and disorders 155
By Dorothea McCarthy

Sex differences in attitudes about sex differences 161
By A. C. Sherriffs and R. F. Jarrett

The intercorrelations between measures of school learning 169
By J. W. Tilton

Manual for scoring the thematic apperception test 181
By Franklyn D. Fry

TAT scoring blank 197
By Franklyn D. Fry

Problem-action responses and emotions in Thematic Apperception Test stories recounted by alcoholic patients 201
By G. A. Knehr, A. Vickery, and M. Guy

A visual imagery technique for the induction of hypnosis in certain refractory subjects 227
By Milton V. Kline

An academic aptitude test for superior college students 229
By H. C. Lindgren, R. Gilberg, and D. W. Crosby

A note on "Familiar size and the perception of depth" 235
By William H. Ittelson

Controlling interaction in small group research 241
By Martin Grossack

Cues, expectations, and first impressions 245
By Martin Grossack

Time-perspective and feeling-tone: A study in the perception of the days 253
By Maurice L. Farber

Simple auditory reaction time in aged adults 259
By Walter D. Obrist

Atypical tonal memory 267
By Cyril C. O'Brien

Methodological considerations underlying electrodermal measurement . . 271
By William W. Grings

Factors influencing routine performance under noise: I. The influence of "set" 283
By E. Victor Mech

Certain factors in producing complexity of response to a single pulse of light 299
By S. Howard Bartley and Frank Ray Wilkinson

Quantitative studies of verbal reinforcement in classroom situations. I. Differential reinforcement related to the frequency of error and correct responses 307
By Donavon Auble and E. Victor Mech

Hypnosis as emotional response: A theoretical approach 313
By Henry Guze

Problem-solving behavior in men with frontal or occipital brain injuries 329
By W. S. Battersby, H. L. Teuber, and M. B. Bender

The performance of albino mice in the maze situation with the utilization of light as motivation and its relative absence as reinforcement . . . 353
By Willard E. Caldwell and H. Lynn Womack

Group memory of meaningful material 361
By Howard V. Perlmutter

Philosophic assumptions in operational psychology 371
By Edward Newbury

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 3-43.

THE RELATION OF OXYGEN DEPRIVATION TO FETAL REFLEX ARCS AND THE DEVELOPMENT OF FETAL BEHAVIOR*[1]

Department of Anatomy, University of Pittsburgh School of Medicine

TRYPHENA HUMPHREY

A. INTRODUCTION

The well known concept of the development of behavior as "the progressive expansion of a perfectly integrated total pattern" was elaborated by Coghill (38, 39) on the basis of his study of amphibian behavior. According to this theory, at the time activity appears it involves all of the neuromuscular apparatus capable of reacting. Subsequently there is "individuation" . . . "of partial patterns which acquire varying degrees of discreteness." Later investigators working on mammals (Windle and his colleagues, for example) have described early local limb movements and believe that isolated limb responses are the earliest fetal reflexes in higher vertebrates. Consequently these observers consider mammalian behavior to be built up by the amalgamation of such isolated reflexes into more complex reactions (148, 149, 150, 151).

In the human embryo, at least, the first recorded response to any known stimulation followed excitation of the maxillary region of the trigeminal nerve of an embryo of 20 mm in crown-rump length (prefixation measurement) and consisted of contralateral flexion of the head (Figure 1), unaccompanied by other movements (54, p. 167). In slightly older fetuses this reaction includes extension of the arms at the shoulders, as a result of pull on the trapezius muscle by the moving trunk, and a rump rotation, also brought about by contraction of the axial muscles (85). Later, according to Hooker (85, 86), the extension of the arms and the rump rotation is caused by active contraction of the limb girdle muscles. Sometimes, although infrequently, the reflex is homolateral (82, 83). Such reactions represent the "total pattern" type of response described by Coghill for Amblystoma. In

---

*Accepted for publication by Leonard Carmichael of the Editorial Board, received in the Editorial Office on August 1, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The physiological and morphological studies on human prenatal development, of which this paper is publication No. 21, have been aided by grants from the Penrose Fund of the American Philosophical Society, from the Carnegie Corporation of New York, from the University of Pittsburgh, and from the Sarah Mellon Scaife Foundation.

the human embryo, as in Amblystoma, responses of this type follow stimulation of nerve fibers transmitting only impulses of exteroceptive nature.

The earliest independent movement of the extremities in the human embryo (Figure 1) in response to known stimulation was noted by Fitzgerald and Windle (54) for an embryo of 23 mm in crown-rump length following stimuli of pressure (by tapping the amnion) and stretch (by flipping the limb), both excitations of proprioceptive nature (161, pp. 606, 612). For the embryo for which arm movements were reported by Yanase (162, stated as 20 mm from crown to coccyx) the nature of the stimulation, whether or not a stimulus was used, the condition of the embryo when observed, and its status when measured (hence its true size) are all unknown. The 22 mm embryo of Strassman (133), measured after fixation (indicating a living embryo of at least 23 mm or more in crown-rump length), showed limb movement which is evidently in response to pressure stimuli for the intact amniotic sac was said to bulge from a laceration in the uterine tube.

In his various papers Windle and his co-workers (54, pp. 162-4; 151, p. 217; 154, p. 93; 156, p. 377; 157, p. 502) have emphasized that the isolated limb responses occur only during the time when the placental circulation remains intact, whereas the contralateral cervical trunk flexion on stimulation of maxillary or mandibular V continues after placental separation (or severance of the umbilical cord) as well (see pp. 17-20). Consequently it is the belief of this group of observers that only the local reactions elicited during the time that the placental circulation is maintained can be considered normal, all others being influenced by anoxia[2] and hence abnormal in character (151). Much has been written concerning these two theories of the development of mammalian fetal behavior without either a reconciliation of the two points of view or a satisfactory explanation of the differences in the observations (see p. 32 and the reviews of Barron, 10, 12, and Windle, 150, 151). Of the various factors which are involved in the elicitation of fetal reflexes, however, some have received relatively little attention.

In the course of studying the spinal tract (or descending root) of the trigeminal nerve and its relation to early fetal activity (92, 93), a considerable amount of the literature dealing with the development of fetal reflexes in both mammalian and submammalian forms was examined. During this time it became increasingly evident that the variations in opinion concerning the manner of development of fetal behavior may have been the result, in part, of failure to recognize or give due consideration to the fact that different types of reflex arcs, anatomically and functionally, are utilized for the cervical flexion on trigeminal stimulation and for the local limb reflex. In other words the "total pattern" type of response is obtained by the use of exteroceptive stimuli (2, 38, 39, 81, 83, 85) and mediated over neuron arcs involving three or more neurons whereas the "local limb reflexes" are secured with proprioceptive stimuli such as muscle stretching (30, p. 118; 32, p. 128; 135, 150, 151 and others) which in the adult are related to two-neuron arcs (see p. 7) or strong pressure (33, 150) which may also produce stretch stimuli (see p. 20).

---

[2] The terms anoxia, anoxemia, and asphyxia appear to be used interchangeably by these observers. According to Wiggers (146, p. 495) anoxemia refers to oxygen deficiency in arterial blood and anoxia is applied to a serious oxygen deficit in the tissues. Asphyxia, however, involves increase in the amount of carbon dioxide and lactic acid as well as oxygen deficiency. The term hypoxia is used for states of oxygen lack that are still well tolerated by the animal.

Since functional capability (as manifested by conduction rate of nerve fibers, threshold of stimulation, and resistance to anoxia, for example) is directly related to the anatomic differences in reflex arcs, and particularly to the type of afferent fibers participating (see p. 12), it seems worthwhile to re-examine the problem of the development of fetal behavior in the light of the types of reflex arcs involved and the effects of oxygen deprivation on the various portions of these reflex arcs. A correlation of certain of these more neglected factors with the differences observed in early fetal reflexes will be attempted in the following pages. For the most part consideration will be given only to the reactions which appear in human fetuses from 7½ to 8½ weeks of menstrual age—the period covered by the author's morphologic study on the spinal tract of V (93).

B. Functional Relations of Different Types of Reflex Arcs

Both two- and three-neuron reflex arcs were described by Cajal (27) and many other early observers. Neurophysiologists usually associate the two-neuron or monosynaptic arcs with the stretch reflexes exclusively (55, 134), whereas visceral reflexes (136) and those of exteroceptive origin (106, p. 117) require at least one, and possibly more internuncial neurons as well. Such reflexes, then, are dependent upon multineuronal or plurisynaptic pathways (89, pp. 94-5; 106, p. 117). Structurally these reflex arcs differ from those consisting of two neurons, the monosynaptic reflexes, in the nature of the fibers which form their afferent limb. Thus the afferent fibers of plurisynaptic reflexes are largely of cutaneous origin and form the delta (general tactile and fast pain) and *C* groups (slow pain; see 88, pp. 794-796 and 863; 146, pp. 116 and 205-207) which vary from myelinated fibers around 10 μ in diameter to the smallest unmyelinated fibers. Fibers of these diameters also include those which mediate thermal sensations. Although some tactile fibers are large in caliber (tactile localization) it is the afferent fibers of less specialized type which are more significant in the skeletal muscle reflexes evoked by exteroceptive stimulation (97, p. 120). Typical of such reflexes is flexion withdrawal on pinching the skin of the foot, a response which is characteristic of the so-called flexion reflexes (55, p. 54).

Anatomical evidence for the presence of two-neuron reflex arcs in mammals has been in existence since 1890 (26), at least, if not longer (see 102) and physiological evidence since 1911 (95). Collaterals of dorsal root fibers passing to the anterior horn cells of the spinal cord were described for human fetuses by both von Lenhossék (102, fetuses 20 to 30 cm long) and von Koelliker (100, fetus of 6 mo.) as well as for the rat, guinea pig, rabbit and newborn mouse by von Lenhossék. In their studies on cat fetuses Windle and his collaborators (147, 160, 161) have called attention to these direct collaterals to ventral horn cells during the earliest stages of fetal activity although Barron (10, 11) did not find that these fibers in sheep fetuses reached the motor neurons until after activity had begun. Such direct collaterals were found by von Lenhossék to be more numerous in the enlargements of the spinal cord. This relationship is likewise demonstrated in Windle's tabulations of the number of collaterals at different cord levels in cat fetuses (147, p. 497; compare C7 with T3).

Von Lenhossék (102) commented on the large caliber of these direct collaterals in the human spinal cord. A large diameter has been demonstrated by physiological methods to be a characteristic feature of the afferent fibers of monosynaptic reflex arcs (see p. 7). In the rodents which he studied von Lenhossék (102, p. 306) even found that these direct collaterals to ventral horn cells pass through the motor cell groups in large part to form a terminal plexus along their outer border. Termination on the ventral horn neurons bordering the white matter suggests a relationship with the nerve cells innervating the extensor muscles, the muscles most closely concerned in monosynaptic reflexes, for according to Bok (18), at least, the extensor muscles are innervated by the superficial cells in the ventral horn.

Recently certain of these anatomical observations on direct root collaterals were confirmed experimentally by Szentagothai (136) by means of bouton degeneration studies after section of the dorsal roots of the spinal nerves. These studies not only demonstrated that certain dorsal root collaterals in the cat pass directly to motor neurons but also that these collaterals are greatest in number in the lower cervical and middle lumbar regions of the spinal cord. In fact they were found to be completely absent only at lowest sacral levels.

Even before the mesencephalic nucleus of the trigeminal nerve was believed to provide the proprioceptive innervation of the muscles of mastication (94), direct collaterals of its root to the motor nucleus of V were identified by Cajal (in 1896 according to Johnson, 94). Since that time these direct connections with the motor neurons have been seen in morphologic preparations by many observers for various mammals including man (139). Thelander (137, cat) first demonstrated these relations with the motor nucleus of V by degeneration methods (Marchi technique). The absence of an interneuron in this reflex arc was demonstrated experimentally by Harrison and Corbin (70). More recently Szentagothai (136) confirmed this direct relationship of the mesencephalic root fibers to the motor trigeminal nucleus by the degeneration of terminal boutons on the motor neurons of V after sectioning its mesencephalic root. In the last study, the degeneration of the large fibers of the annulo-spiral endings and, less often, the medium sized fibers of flower-spray endings in muscle spindles of masticatory muscles demonstrates clearly that internuncials are lacking in this reflex arc. Since Corbin and Harrison (45) observed electrical activity in the mesencephalic tract of V on lowering the mandible (stretching the masticatory muscles) the nature of this two-neuron reflex arc as a stretch reflex seems clearly established.

It has been suggested by various observers that all two-neuron, or monosynaptic reflex arcs are proprioceptive in nature and that the afferent limb of such an arc carries impulses set up only by stretch stimuli (106, 107, 109, 136). As indicated by variations in the transmission time for different reflexes through the spinal cord, there is at least little doubt of the existence of monosynaptic reflex arcs and their close relation to stretch reflexes (106, 109, 111, 125; see also 89, p. 94 and 146, p. 155). Observations on the rate of conduction of afferent impulses from muscles (146, p. 205; 88, p. 796) have shown these stimuli to be transmitted by the large, heavily myelinated fibers which lie in the upper third of the fiber spectrum (12-20 μ; 146, p. 115) and which conduct most rapidly (115-120 meters per second). Consequently the shorter conduction time and briefer interval for synapse (the central latency period) following stimulation of afferent fibers by stretching their related muscles, as compared with the greater conduction time found on excitation of cutaneous nerves, clearly indicates that stretch stimuli are associated with two-neuron or monosynaptic reflexes (106, 108).

Stretch reflexes are often divided into two classes—the phasic and static stretch reflexes (15, p. 949; 89, p. 84; 146, pp. 154-5). According to Liddell and Sherrington (103, p. 223) and Creed et al. (47, p. 50), however, all stretch reflexes have both a phasic or kinetic component and a static or postural (tonic) component. The former results from stretch movement (a sudden pull) and lasts only a fraction of a second but is intense in nature; the latter is evoked in response to a passively imposed posture and lasts much longer but is less powerful. Tendon reflexes of all muscles and clonus (such as patellar clonus) represent one fraction of the complete stretch reflex—the phasic element (103, p. 233). The other constituent is tonic or postural in nature and constitutes the static aspect of the stretch reflex. The physiologic differences in these two components of the complete stretch reflex are emphasized by decerebration, which accentuates the postural element, and by spinal cord section, which practically abolishes it (103, p. 233).

In spite of the fact that both components of the stretch reflex are better developed in the extensor or antigravity muscles than in the flexors, stretch reflexes have been demonstrated in flexors by many investigators (4, p. 278; 117, p, 27; 113, p. 59). According to Howell (89, p. 85) all muscles which have been examined "respond in some degree to the stimulus of stretch" (see also 47, p. 60; 78, p. 11; 79, p. 26; 103, p. 236; 107, p. 312; 109, p. 4). Even facial muscles such as the orbicularis oris are considered by some observers to exhibit a myotatic reflex (52, p. 110). It is the tendon jerk or phasic component of the stretch reflex, however, which is best represented in the flexor muscles whereas the static element may be almost completely absent (103, p. 236). Unfortunately the use of the term "pluck reflex" for the tendon reflexes in response to stretch stimuli of flexor muscles (55, p. 127; 107, p. 312) tends to obscure the fact that the tendon jerks of flexors, as well as those of extensor muscles, are stretch reflexes mediated by two-neuron reflex arcs (107, p. 312; 109, p. 4). Nevertheless the nature of these tendon jerks as stretch reflexes of flexor muscles has been commented upon by various observers (47, p. 60; 78, p. 6-9; 103, p. 236; 107, pp. 310-312; 109, p. 4) and in the recent study of Eccles and McIntyre (51, p. 466) on plasticity of mammalian monosynaptic reflexes, responses were examined for both flexors and extensors.

Stretch reflexes were designated as myotatic reflexes by Liddell and Sherrington (103, p. 240; 104, p. 273), who applied the term to all reflexes in response to stretch, regardless of whether they occur in flexor or extensor muscles or are primarily phasic or static in nature. Evidently the term "myotatic reflex" is used in a narrower sense by some observers who apparently reserve it, at times, for stretch responses in extensor muscles exclusively (107, p. 312, Table 2), although clearly stating, not only that flexor tendon jerks are "two-neuron-arc reflexes" (p. 312), but also that "the two-neuron-arc pathways are reserved for mediation of the myotatic reflex" (p. 313).

It should be pointed out that even in the adult, where presumably they might be combined with each other if behavior is built up by the amalgamation of unit reflexes (see p. 20) the stretch reflex itself remains localized in nature. Actually Sherrington (127, p. 86) was one of the earlier observers to note this characteristic of the phasic stretch reflexes. Others have demonstrated that these responses are so restricted in nature that only part of a muscle responds if only that portion is stretched (44, 103, 104, 107, 108, 111; see also 89, pp. 86-7; 146, p. 155). This local character of the phasic stretch or tendon reflex is further evidenced by the fact that only strictly limited spinal cord segments can be shown to be intact by elicitation of any single tendon reflex (for example the biceps jerk, C5-6, or triceps jerk, C6-7; 15, p. 949). Although these phasic stretch or tendon reflexes, as well as the postural or static stretch reflexes, are influenced (i.e., inhibited or facilitated) by higher nervous system centers such as the reticular formation, vestibular nuclei and cerebellum (116, 132), the individual tendon reflex itself is always a localized reaction even in the adult. It is only indirectly, that is by the pull of the contracting muscle on its antagonist, that a stretch stimulus acts on an adjacent joint in addition to the one it crosses, although such a stretch stimulus does facilitate directly the synergists of the muscle stretched (111). The effect of higher nervous system centers, then, does not appear to be so much that of integrating these local stretch reflexes into a pattern with other reflexes, as has been suggested by Windle (see p. 32), but of their facilitation, or inhibition, and the stabilization of the muscle tonus itself.

These physiologic observations on the local character of phasic stretch reflexes are in harmony with the early anatomical studies of von Lenhossék (102, p. 304) who found the direct collaterals to ventral horn cells to be given off mostly at the level of entrance of the nerve roots. Recently the degeneration studies of Szentagothai (136) have confirmed these anatomic findings and in addition have shown that these direct collaterals are not only given off mainly at the level of entrance of the nerve, but also show a sharp decrease in number immediately above and below this level and a limit of not over one and one-half segments in either direction.

In the foregoing discussion of monosynaptic and plurisynaptic (or multisynaptic) reflex arcs, consideration has been given primarily to the neuronal pathways for reflex activation of the agonist muscles, those which produce the movement in question. That "an inhibitory process is part and parcel of the reflex reaction" was early recognized by Sherrington (127, p. 90) and has been demonstrated repeatedly experimentally (see 47 and 55). Thus even the primary two- and three-neuron reflexes under discussion in this paper undoubtedly involve additional neuronal relations, for the antagonistic muscles, if capable of contracting, must be inhibited at the same time that the contracting muscles (the primary movers, or agonists, and the synergists) are excited.

In addition to the sensory root fibers which pass directly to motor neurons for excitation of the agonist (or primary muscle which contracts), some observers (89, p. 95; 105, p. 190; 111, p. 442; 136, p. 453) believe that collateral branches of these dorsal roots also pass to the motor neurons supplying the antagonistic muscles at the same joint to function in an inhibitory capacity. On the basis of physiological evidence Lloyd (111, p. 442) suggests that both the inhibition of the antagonists and the facilitation of the synergistic muscles (i.e., those acting on the same joint as the agonist excited by stretch stimuli) function through two-neuron reflex arcs. This concept is supported by the fact that Szentagothai (136, p. 450) found some bouton degeneration in relation with the motor neurons supplying the antagonists of the masticatory muscles as well as more extensive degeneration on the motor neurons supplying the agonists (see p. 7).

During recent years various other mechanisms whereby reflexes may be inhibited have received attention (see 14, 16, 21, 22, 23, 24, 58, 61, 101, 110, 111, 113, 114, 116, 119, 126). In addition the subject of central inhibition is reviewed by Creed et al. (47) and by Fulton (55). Although several theories for explaining inhibition have been presented, only certain aspects of the problem appear pertinent to the discussion at hand. One of these is the fact that a phasic stretch or tendon reflex, like the knee jerk, is less resistant to the onset of asphyxia[3] than is the inhibition of this reflex (68, p. 101). This observation is in harmony with the belief of Matthews (117), Granit (61) and McCouch, Deering, and Stewart (119) that the tendon spindles of

---

[3] The functional survival of individual neurons on exposure to anoxia (as demonstrated by their ability to transmit nervous impulses) differs from their capacity to remain alive after prolonged periods of anoxia. Thus various observers (67, 68, 69, 75) have demonstrated that after periods of asphyxia during which all reflexes remain suppressed for varying intervals, the tendon reflexes may be able to resume function (or resume function after a shorter time lapse) whereas the exteroceptively elicited reflexes may remain permanently abolished (or return only after a longer interval). According to Gerard (57, p. 988) the interval during which the neuron itself can survive anoxia is dependent upon such factors as proteolysis and the accumulation of acid and need not necessarily be related to the survival time of its function although it may parallel it (see p. 15). As a result, although the large sensory neurons with a high metabolic rate (20, p. 357; 146, p. 116), lose the ability to respond to stimulation under the influence of anoxia sooner than do the small ones with lower rates of metabolism, these large ganglion cells are able to resume function after longer periods of anoxia.

Golgi are inhibitory in function, since these end organs are associated with somewhat smaller fibers, which are probably more resistant to the onset of anoxia than are the large fibers mediating stretch reflexes (see p. 12). Although the work of van Harreveld (67, 68, 69) has shown that both tendon reflexes (stretch stimuli) and flexion reflexes (exteroceptive stimuli) survive prolonged periods of asphyxia which permanently suppress their inhibition, no work has been found which demonstrates whether or not exteroceptive reflexes, as well as tendon reflexes, are less resistant to the onset of anoxia than is their inhibition. Inasmuch as it is only the elicitation of the reflex in question, when under the influence of anoxia, that is considered here, a knowledge of these relationships is not pertinent to the discussion.

It should be noted also that the present consideration of reflex arcs is confined to those mediating reflexes which require only the sensory and motor limbs of a reflex arc for their execution, together with an internuncial neuron when one is present. Activity which involves the interpolation of additional neuronal relations, even those of the reticulo-spinal pathways for instance, should not be confused with these primary reflexes, referred to by Sherrington (127, p. 320) as "fundamental" reflexes since "all the other neural arcs" are superimposed on them. Thus the following discussion is not intended to apply to the more complex reactions which necessitate the functioning of integrative centers of the midbrain, striatum, or other areas of the brain. In order to avoid confusion of these more complex reactions with the fundamental or primary reflex arcs, the former will be referred to as secondary and the latter as primary reflexes.

C. The Effects of Oxygen Deprivation on Reflex Arcs in the Adult

The fact that severe oxygen deprivation abolishes reflexes has been known for many years (127, p. 79). That a lesser degree of oxygen want may be the cause of increased excitability was suggested by Sherrington (128, p. 220) who stated that "a certain degree of asphyxia much favors the elicitation of the scratch-reflex." Even earlier Kaya and Starling (98, p. 353) had noted that "sudden excitation of the spinal centres" appeared with lower oxygen tension. In these early observations the manner in which the different portions of the reflex arcs were affected by the lack of oxygen was neither determined nor suggested.

The selective effects of oxygen deprivation on reflex arcs have been demonstrated experimentally (20, 99, 123). On the motor side of the primary reflex arc Groat and Koenig (63) and Lloyd (112) have even determined the effect of asphyxia on various portions of the primary motor neuron. On the sensory side of these arcs a differential susceptibility of the presynaptic fibers has also been demonstrated. On the one hand, it has been shown that the large proprioceptive afferent fibers which are associated with the stretch reflexes not only have a low threshold for stimulation (106, p. 116) but also are among the earliest fibers to be affected by and to succumb to the influence of asphyxia[4] (88, p. 863; 146, p. 116). Hence on exposure to oxygen deprivation proprioceptive reflexes should be the earliest to be facilitated initially (68) and either appear, if they have been absent previously, or, if already present, become more readily elicited and/or greater in amplitude. As anoxia progresses to the point where blocking of the reflex occurs, however, these reflexes should also be the earliest of all of the primary reflexes to be suppressed (68). On the other hand, the smaller exteroceptive presynaptic fibers which are associated with multisynaptic reflex arcs have a higher threshold for stimulation (15, p. 914; 88, p. 794) and succumb to asphyxia later than do the proprioceptive fibers (65, 88, p. 863; 146, p. 116). Consequently exteroceptive reflexes dependent upon these fibers should not only be facilitated by anoxia later than are the proprioceptive reflexes but also should be blocked only after a relatively longer period of exposure to oxygen want. Inasmuch as the slow pain fibers are the last of all of the presynaptic fibers to be blocked by anoxia (88, p. 863; 146, p. 116) reflexes related to these fibers should be the last of the skeletal muscle reflexes to be affected by anoxia or asphyxia.

Concerning the primary reflex arcs as functional units Brooks and Eccles (20, p. 353) have demonstrated for monosynaptic reflexes that in asphyxia "the blockage of the reflex pathway is initially due to failure of propagation in the presynaptic fibers." As summarized by these authors (20, p. 358): "The presynaptic terminals of the afferent fibers undergo a change first, as indicated by the early increase in amplitude and subsequent disappearance of their potentials." These neurophysiologists state further that the cell bodies of the motor neurons have the "next highest oxygen requirement" and that the period of hyperexcitability of these cell bodies "outlasts that of the presynaptic fibers but they in turn fail before either the motor axons or the more peripheral part of the afferent" fibers. Thus the blocking of the monosynaptic reflexes, at least, is dependent upon the fibers which form the afferent limb of the reflex arc.

---

[4] According to Brooks and Eccles (20, p. 354): "When animals are subjected to oxygen lack uncomplicated by excess of carbon dioxide, the results obtained do not in the main differ from those obtained in asphyxia. The hyperexcitability phenomena develop gradually without any sign of the initial depression sometimes observed with asphyxia." In further discussion these authors add (20, pp. 354-5): "Continuation of oxygen lack eventuates in the typical depression of all potentials just as in asphyxia. As would be expected, the rate of failure depends upon the amount of oxygen in the gas mixtures given." Thus it is evident that, for a discussion of the effects of oxygen deprivation on fetal behavior, it matters little if at all whether anoxia or asphyxia is being considered.

It should be pointed out here that these observations of Brooks and Eccles (20) are based on animals anesthetized with nembutal. Such barbiturates block the internuncials of the plurisynaptic reflex arcs at a depth of anesthesia which permits the monosynaptic reflexes to remain functional under experimental conditions (19, 20, 25, 50, 66), although greater degrees of anesthesia block the monosynaptic reflexes as well. Consequently these experiments do not indicate whether it is the presynaptic sensory fibers of the plurisynaptic reflex arcs or the internuncials which fail under the influence of anoxia before the motor neurons are blocked. Nevertheless these investigators do state specifically that the point of reflex blockage in asphyxiation differs from that with nembutal in that "blockage of the reflex pathway occurs initially in the presynaptic fibers" (20, p. 357). Moreover, van Harreveld (68, p. 101) has demonstrated that the proprioceptively elicited knee jerk succumbs to asphyxia earlier than the exteroceptively evoked flexion reflex (in 81 per cent of 48 trials using 9 spinal cats under light nembutal anesthesia; 68, pp. 98-99 and Figure 1A and B and Table 1). It may be concluded, then, that the relative resistance of primary reflexes to the influence of anoxia or asphyxia is dependent upon the afferent limb of the primary reflex arcs, at least in large part.

It might be suggested that the components of the primary reflex arcs are affected by anoxia in the reverse order of their functional development, that is, the presynaptic fibers first, in a sequence related to their size and degree of specialization so that the most highly differentiated fibers are blocked earliest of all, the internuncial neurons following the presynaptic fibers, and the primary motor neurons last of all (see Hooker, 85, for a discussion of the developmental relationships in submammalian vertebrates). Such a relationship would indicate an oxygen gradient within the primary reflex arcs comparable to that of the central nervous system as a whole (see p. 14) since here too the latest developing regions are the earliest to succumb on exposure to anoxia. In any case it is evident that the blockage of primary reflexes under the influence of anoxia is dependent in major part, if not entirely, upon whether the presynaptic fibers concerned are the large proprioceptive afferents or the smaller fibers functioning in an exteroceptive capacity.

Because the effects of anesthesia on the transmission of nervous impulses may be of significance in interpreting mammalian fetal activity studies (83, 85, 150, 151), some mention of these relations should be made here as well. The various studies reported by Himwich (75) indicate that the anesthetic effects of nitrous oxide are manifested only when the oxygen supply is markedly reduced. Hence the influence of nitrous oxide on the nervous system appears to be largely that of anoxia. In ether anesthesia likewise (124) decreased oxygen consumption occurs. Evidently the significant effect of ether and chloroform on primary reflexes resembles that of anoxia and asphyxia for these anesthetics not only block the presynaptic fibers but also suppress the higher integrative motor centers before the lower motor neurons are inactivated (48, p. 273; 64, p. 47). In spinal blocks with cocaine (56, p. 55), although the unmedullated *C* fibers succumb very early the medullated fibers then follow in the order of the degree of myelination (or fiber size) with those having the largest myelin sheaths blocked first. As a result the large proprioceptive afferents should be inactivated before the impulses traveling the smaller fast pain and general tactile fibers are suppressed. The work of Kato (97) indicates that the large afferent fibers with low threshold also succumb to the effects of some narcotics (urethane) before the smaller high threshold fibers fail to function.

Although the oxygen consumption of the cerebral cortex is decreased more than that of lower centers when under the influence of barbiturates (see discussion in 124), these anesthetics also block the internuncials of the primary reflex arcs before inactivating the large presynaptic fibers functioning in monosynaptic reflexes (66, see also p. 13). This observation is in harmony with those of Hooker on human fetal activity, for this observer (82, p. 200) found that "the barbiturates and avertin more or less completely suppress fetal activity during the period of their effectiveness." Inasmuch as Hooker does not base his reports on fetuses under the influence of such anesthetics (83, 85, 86) the differential effects of barbiturates on the two types of primary reflexes really have no bearing on the controversial aspects of the early development of fetal behavior.

The effects of anoxia (or asphyxia, see footnote on p. 12) on the primary reflexes should not be confused with their action on the central nervous system as a whole. As is well known, a complete anoxia can be tolerated by the adult central nervous system for only a short time (120, p. 335; 146, p. 502; 153). Moreover, the functions of the higher nervous system centers are lost under the influence of oxygen deprivation much more quickly than are those of lower centers such as the medulla oblongata (57, p. 988). According to Cannon and Burket (28), the maximum survival time for the spinal cord under asphyxial conditions is 60 minutes as compared to 8 minutes for the cerebrum.

More recent observers (57, 62, 72, 75) report a similar gradient of resistance to nerve cell injury by asphyxia, that is, neurons of the cerebral cortex are damaged earliest, then those of the cerebellar cortex, basal ganglia, and sensory and integrative centers of the midbrain, pons, and medulla in that order. Others have found cerebellar function to fail under anoxial conditions even earlier than that of the cerebrum (57, p. 987). However, the neurons which constitute the primary reflex arcs show the greatest resistance to anoxia, both in regard to loss of function as anoxia progresses and to recovery after exposure to intervals of anoxia (57, p. 988). Consequently primary reflexes are retained after higher integrative nervous system centers have succumbed to anoxia.

The evidence available, moreover, does not warrant the assumption that these primary reflexes themselves become abnormal in character before they disappear, as the discussions of Windle (151 and elsewhere) suggest. On the contrary, the studies of Porter, Blair, and Bohmfalk (123) and Brooks and Eccles (20) indicate that stretch reflexes depending upon monosynaptic reflex arcs retain their physiological characteristics so long as they can be elicited at all. Actually, upon a decrease in oxygen there is initially facilitation, manifested by a greater amplitude of contraction (20), due to the inclusion of more motor neurons in the response than are activated with a normal oxygen supply (60, p. 566; 97, p. 44; 123, p. 167). This increase in amplitude is followed, as oxygen deprivation continues, by a decrease in amplitude of contraction and eventually by a failure of the reflex response (20, p. 357) as progressively more neurons succumb to anoxia. The figures of such observers as Brooks and Eccles (20, Figures 2-3) indicate, in addition, delay in the time between stimulation and response, but the characteristics of the tracing produced by the monosynaptic reflex are retained. These observers noted similar effects resulting from anesthesia. In other words the monosynaptic reflex, at least, remains typical under anoxia (and anesthesia) so long as it can be elicited. Such reflexes cannot be considered to become abnormal *per se*—unless variations in amplitude and synaptic delay constitute abnormalities—nor do they take on the peculiarities of exteroceptive reflexes.

From a consideration of the gradient of resistance of the central nervous system to anoxia, however, it will be evident that, after higher nervous system centers become functional, as in the adult animal (or in fetal stages later than those under consideration here, see below), the nature of the activity observed will be altered by a progressing anoxia (or asphyxia). Thus, in anoxia, behavior will proceed from what is normal for the adult or older fetus at the age in question—that dependent upon cortical, striatal, midbrain or other higher nervous system centers as the case may be—to types of activity which are mediated at progressively lower levels of the central nervous system. Finally only the primary reflexes dependent upon the motor and sensory neurons (and internuncials) will remain. Obviously such behavior is of abnormal or pathologic character for the normal adult or for the older fetus, inasmuch as such individuals are capable of activity dependent on higher levels of the nervous system. However, for the young human fetus of 7½ to 8½ weeks of menstrual age which is able to perform only the primary reflexes, any reflex elicited will not only conform to its type but also will be characteristic for the age period.

D. The Effects of Oxygen Deprivation on Fetal Reflexes

In addition to demonstrating that the fetal and newborn mammal is more resistant to anoxia than is the adult (53, 75, 152, p. 54), the available studies also show that the younger fetuses survive longer than do the older ones (59, see discussion in 75, pp. 129-130). These observations are in accord with the work of Barcroft (5) and of Boell and Nicholas (17). Barcroft states that the fetus uses only about 10 per cent of its available oxygen during early embryonic life although in late fetal life it utilizes nearly all of it (5, p. 9). As a result the younger fetus on delivery has a greater oxygen reserve to be used, before suffering the effects of anoxia, than does the older one. Moreover, because the oxygen uptake per hour is greater in older fetuses (17) even the already depleted oxygen reserve will be used up more rapidly in older fetuses following delivery.

Although the time of survival under exposure to anoxia is greater for the fetus than for the adult, the numerous studies of mammalian fetal activity offer ample evidence that the fetal nervous system responds to anoxia in essentially the same manner as does that of the adult (see p. 11). Thus oxygen want leads to increased excitability followed by complete blocking of the response as the anoxia progresses (see 150) even though differentiation of the neurons is not complete at the time fetal activity begins (no myelin has been demonstrated and both receptors and effectors are still undifferentiated; see also p. 26). The excitatory, or better the facilitatory, effect (123, p. 171) of anoxia on reflexes has been commented upon by various observers of fetal behavior (1, 6, pp. 330-331; 10, p. 11; 150, p. 251; 151, pp. 220-221; 154, p. 94; 156, pp. 377-379). Likewise, the retarding effect of increased anoxia or of asphyxia has been mentioned by these investigators as well as by other workers in this field. The subject is discussed in some detail in the papers of Barron (10) and Windle (150, 151). However, these discourses have not dealt with the relative susceptibility to anoxia of the different parts of the reflex arc itself and, in turn, the comparative sensitivity to oxygen deprivation of the different types of sensory or presynaptic fibers which vary in their reaction to oxygen deprivation and succumb earlier than do the motor neurons.

An examination of the various reports on fetal behavior indicates that the local limb movements produced by Swenson (135) and Windle and his collaborators (see 54, 159, 161) on flipping the forelimb of the fetus, if of reflex character as these authors believe (see 150), are of the myotatic or stretch reflex type (broken lines, Figure 1). That this is true is suggested in part by the fact that these movements, as described by Windle and his colleagues (148, p. 641; 154, p. 93) and others (8, p. 491; 10, p. 17), involve particularly the extensor and abductor muscles—the muscles most active in the stretch reflexes in the adult (see p. 000). Further evidence that the local limb movements of the fetus may be stretch or myotatic reflexes (if they are reflexes and not the result of direct muscle stimulation as they have been considered by some observers of fetal activity; for example, 10, 43, 83) is their short latent period following the application of the stimulus and the extremely rapid character of the response (161, p. 605), both qualities which are also characteristic of stretch reflexes (106, p. 117).

The effect of anoxia on the elicitation of the local limb responses of the fetus is also characteristic of the stretch reflexes. Thus Windle and his coworkers (54, 156, 158) have stressed repeatedly that these local limb movements can be evoked only shortly after delivery of the fetus and during the time that the placental circulation is intact. They have indicated further that a "partial anoxemia" not only facilitates the appearance of the limb responses (156, p. 377) but that in the fetus this effect apparently lasts longer than in the adult (p. 379). Hence, on exposure of the fetus in utero, with the placenta intact, these workers report that as the blood began to darken in the umbilical veins "the fetus started to respond to stimulation" (156, p. 377). These authors have also pointed out that even minor degrees of anoxia are enough to prevent the elicitation of the local limb movements (54, p. 163; 154, p. 93; 156, p. 377) and in his 1950 paper Windle (151, p. 219) laid great stress upon the ease with which the early limb activity is abolished under the influence of "anoxemia" alone (see footnote, p. 4).

According to Windle and his collaborators (54, p. 162; 147, p. 488; 151, p. 216) local reflexes are also produced by pressure stimuli such as direct pressure on the muscle or tapping on the amnion. These investigators (151, pp. 217, 221; 161, pp. 606, 612) and others (7, p. 113; 8, p. 491; 10, p. 18)

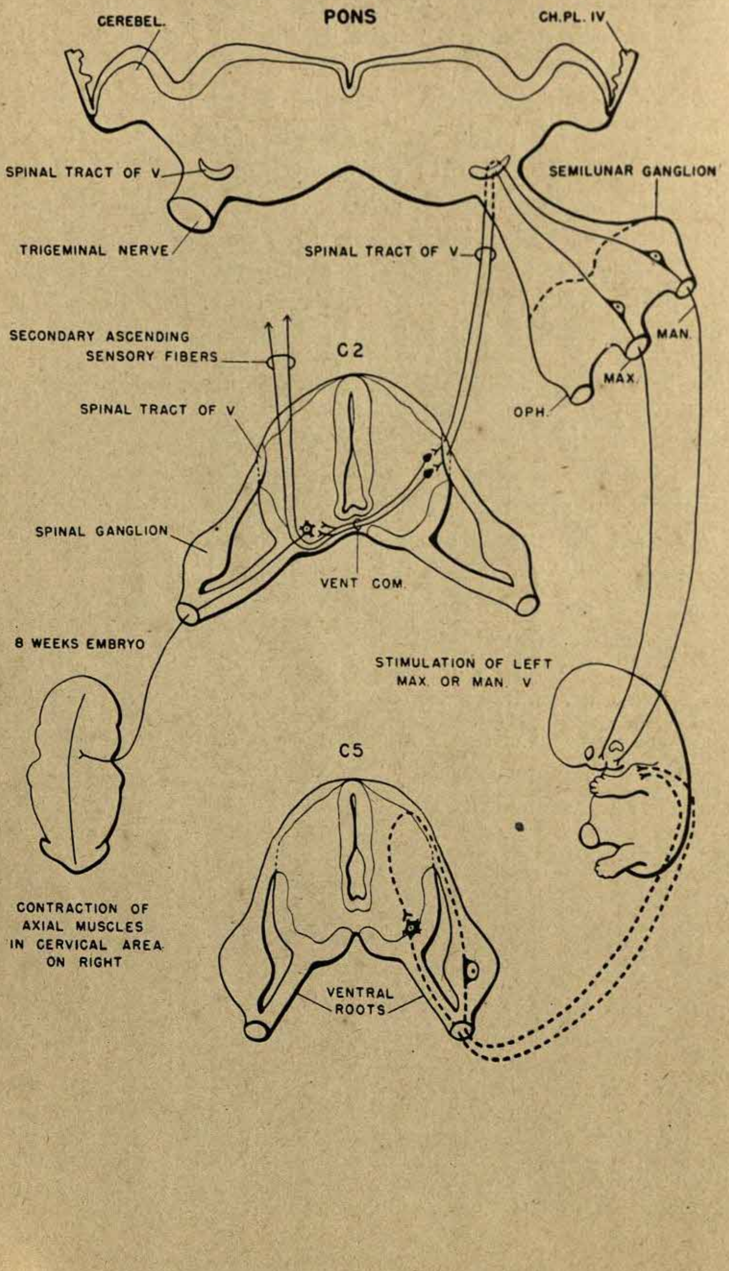

CEREBEL.
PONS
CH.PL. IV
SPINAL TRACT OF V
SEMILUNAR GANGLION
TRIGEMINAL NERVE
SPINAL TRACT OF V
SECONDARY ASCENDING SENSORY FIBERS
C2
SPINAL TRACT OF V
MAN.
MAX.
OPH.
SPINAL GANGLION
VENT COM.
8 WEEKS EMBRYO
STIMULATION OF LEFT MAX. OR MAN. V
C5
CONTRACTION OF AXIAL MUSCLES IN CERVICAL AREA ON RIGHT
VENTRAL ROOTS

FIGURE 1

A DIAGRAM TO ILLUSTRATE THE DIFFERENCES IN THE REFLEX PATHWAYS FOR EXTEROCEPTIVE AND PROPRIOCEPTIVE TYPES OF STIMULATION

Solid lines are used to show the course of the nervous impulses set up on stimulation of the maxillary or mandibular divisions of the trigeminal nerve. The area of stimulation is shown by the 8-weeks fetus on the right and the contralateral trunk flexion in the cervical region is indicated by the fetus on the left. This part of the diagram is based on the author's findings on the extent of the trigeminal nerve fibers in the spinal cord as illustrated in a previous paper (93). Note that the exteroceptive reflex arc involves a third neuron or internuncial (plurisynaptic or multisynaptic reflex), which is illustrated here as a collateral off of the ascending sensory fibers. The broken lines show the course of impulses for a proprioceptive reflex in response to stretch stimulation of such a muscle as the deltoid. A reflex of this type is localized to the spinal cord segments innervating the deltoid and has been demonstrated for the adult to involve only the afferent and efferent neurons of a reflex arc without necessarily involving the internuncials (see p. 28). It is this class of reflex arc which is probably involved in the responses to flipping a limb, direct pressure (if the muscle itself is not stimulated), or tapping on the amnion (producing pressure and so stretching muscles).

The sections of the spinal cord and pons, as well as those of the sensory ganglia, are drawn to the same magnification from a pyridine silver preparation of a fetus of 8 weeks (Homo 131). *C2 and C5,* second and fifth cervical segments of the spinal cord respectively; *CEREBEL.,* cerebellum; *CH.PL.IV,* choroid plexus of fourth ventricle; *MAX.,* maxillary division of trigeminal nerve; *MAN.,* mandibular division of trigeminal nerve; *OPH.,* ophthalmic division of trigeminal nerve; *V,* trigeminal nerve; *VENT.COM.,* ventral commissure.

consider such pressure stimuli, as well as stretching the muscles, to be proprioceptive in character. Examination of the limb movements which appear in response to pressure stimuli demonstrates that they too may be stretch reflexes resulting from the production of tension in the muscles. At least Hoffmann (78, p. 3) found that pressure on a muscle itself as well as a tap on a tendon (p. 7) produces a momentary stretch stimulus which is adequate to elicit a muscle contraction in the adult. Hoffmann (78, p. 9) estimated a pull of 1/5 mm necessary for the human biceps muscle in the living subject. Stretch of 0.05 mm, or less, in 1/20 of a second may be sufficient to produce a response according to Liddell and Sherrington (103, 106), and Denny-Brown and Liddell (49) found an extension of only 8 μ necessary, after decerebration, to elicit a stretch reflex in muscle with fibers 40 mm long. Although it is manifestly impossible to calculate the strength of the pressure stimulus applied when the intact amnion is tapped or a muscle is pinched, such a stimulus is obviously more than sufficient to provide the minute amount of stretch necessary to stimulate the afferent fibers of the myotatic reflex arcs of a young fetus if such reflex arcs are capable of functioning.

No experimental physiological work on fetal stretch reflexes, such as has been done on the adult, has been found in the literature. Consequently it is not possible to say whether stretch reflexes are more or less marked in extensors as compared with flexors in fetal life. Inasmuch as all muscles are slow in contraction in the newborn animal according to Creed et al. (47, p. 60) and the rapid muscles acquire speed in contraction only during the later course of development, stretch reflexes may be relatively more active in the flexors of fetuses than they have been found to be in the adult, provided, of course, that the requisite amount of oxygen is available.

It should be noted that stretch reflexes, even in the adult, are remarkably localized in character (see p. 9). As there is every indication that the local limb reactions of the fetus are proprioceptive stretch reflexes, if they are reflexes, it seems highly unlikely that complex behavior would be built up simply by their amalgamation into an intricate pattern, as Windle believes (151 and earlier). It is more probable that the local limb reflex, from the time of its inception in fetal life, involves all of the neuromuscular mechanism which is able to respond to the stimulus in question just as does the adult stretch reflex. If fetal stretch reflexes change during development, it is more likely that the degree of localization of the response is heightened as the muscle spindles differentiate than that the individual reflexes amalgamate. The apparent increase in complexity which Windle (151 and earlier) considers to be the basis for the integration of behavior may be due in part to concomitant exteroceptive and proprioceptive excitation as well as to simultaneous or sequential stimulation (32, 87) of other muscles in a limb as the contraction of one muscle stretches another (89, p. 84; 111, p. 442). Hence these more complicated reactions occur more often when a greater portion of the cutaneous surface can be stimulated and more of the motor mechanism is able to respond, as intimated by Barcroft and Barron (8, p. 491). At least the integration of localized reflexes into the pattern of activity is not secondarily accomplished after they can be elicited but is an essential part of their initial appearance (see p. 32).

The fetal reflexes produced by stroking the skin with a horse hair (2), a brush (46), a calibrated hair "esthesiometer" (31), human and horse hairs (81, 82, 83, 85) or sometimes by "strong mechanical stimulation" of the snout (54, p. 167), like the responses following cutaneous stimulation in the adult, are exteroceptive reflexes mediated by plurisynaptic reflex arcs (solid lines, Figure 1). As in the adult the fetal reflexes elicited by such stimuli are more resistant to anoxia than are the local limb reflexes dependent on proprioceptive stimuli (151, p. 216). According to Windle and his colleagues these exteroceptive reflexes may occur later after the delivery of the fetus than do the local limb reflexes, although they may also appear during the time when the limb reflexes can be elicited (154, p. 93; 161, p. 611) and are retained for some time after the limb responses have disappeared entirely (158, p. 164; 151, p. 219; 154, p. 93; 156, p. 379). Inability to demonstrate these exteroceptive reflexes immediately on delivery of the fetus, but only after the onset of oxygn deprivation, indicates that they, too, like the stretch reflexes, are facilitated initially by oxygen want, the essential difference being the higher level of anoxia needed. As anoxia progresses, however, Hooker (81 et seq.) has found it necessary to use increasingly stronger stimuli to elicit these reflexes. Such observations denote the blocking effects of a greater degree of anoxia. Eventually these plurisynaptic exteroceptive reflexes are also abolished by anoxia just as are such reflexes in the adult.

It is evident, then, that in fetal development both the local limb movements produced by stretch stimuli and the contralateral axial flexion resulting from cutaneous stimulation are facilitated initially by a slight degree of oxygen deprivation, are retarded as anoxia advances, and are blocked completely when oxygen deprivation becomes still more severe. The essential difference between the reaction of these two types of reflexes to anoxia is the threshold level at which the presynaptic fibers constituting the afferent limb of the reflex arc are influenced by the oxygen want. If the reflexes

11-8-70
9111

resulting from exteroceptive trigeminal stimulation with their higher threshold for reaction to anoxia are to be considered abnormal because of the effects of oxygen deprivation, as Windle believes (151, see also p. 4), then those which follow proprioceptive stretch stimuli, where the threshold for reaction to anoxia is lower, must be at least equally abnormal in character. Since the fetal reflexes obviously respond to anoxia in the same manner as do those of the adult, both the proprioceptive stretch reflexes and the exteroceptive plurisynaptic reflexes are undoubtedly merely increased or decreased in amplitude and facilitated or delayed, depending upon the degree of anoxia (or asphyxia) so long as they are not suppressed entirely. Primary reflexes of fetuses under the influence of some anesthetics as well probably remain true to type so long as they can be elicited (see p. 15).

Although the condition of the fetus in amnio approaches the fetal relations in the intact uterus of a normal pregnancy more nearly than does its status after placental separation, as stressed by Windle (150, 151 and earlier), even the trauma incident to manipulation of the uterus, its exposure to room temperature and the opening of its wall, is probably sufficient to bring about reflex constriction of the uterine blood vessels and hence alter the blood supply to the placenta (86, p. 191). Up to the present time it has been impossible to determine whether all of the fetal movements which have been observed in mammals take place in the intact uterus during pregnancy. The types of stimuli which are present in utero are probably somewhat different in character and are unquestionably fewer in number than after delivery, as suggested by Windle (149, p. 165; 150, p. 258) and demonstrated by transillumination experiments for the sheep by Barron (10). Pressure of other viscera against the uterus and perhaps entanglement of a limb with the umbilical cord might provide stimulation for stretch reflexes. Probably contact with the amniotic wall produces a certain amount of exteroceptive stimulation as well. Also some decrease in the degree of oxygenation of the fetal blood, perhaps sufficient to facilitate reflex responses, may occur as it does in the adult. Indeed, the validity of such an assumption has been demonstrated by Windle (149, p. 167) for the cat, for this observer found increased activity on stimulation of fetuses through the abdominal wall if "anoxemia" was induced in the mother.

Regardless of the fact that it is impossible to state whether the specific mammalian fetal reflexes which have been observed after opening the uterus also occur during pregnancy in the same sequence, at least the fact remains that the neuromuscular system is capable of producing these reflexes at the ages in question could the adequate stimulus be supplied. That such is the case is indicated by the experiments of Harrison (71), Carmichael (29), and Matthews and Detwiler (118) who found that amphibian larvae developed in an anesthetic solution which prevented reflex responses executed normal swimming movements in a few minutes after transfer to fresh water (see also 32 and 85). It is unlikely, therefore, either that the responses noted in mammalian fetal activity studies are not a normal part of the fetal repertoire or that the sequence for the early development of reflex responses in utero differs in any essential feature from that which has been observed thus far. In view of the fact that stimuli in utero are probably few and that the oxygen requirement during early embryonic life is relatively low (5, p. 9; 152, p. 33), ordinarily there will be no oxygen deprivation during early fetal life to facilitate reflexes. In addition, when reflexes are first elicited the threshold for stimulation may be high (46, p. 368; 149, p. 148; 160, p. 530). Consequently responses which have been observed after delivery of the youngest fetuses (or in amnio with the placental circulation intact) probably do not take place normally in utero until a slightly later developmental age (151, p. 221) than that at which they have been demonstrated in fetal behavior studies. It is highly unlikely, however, that movements with so small an amplitude as those observed after delivery of human fetuses of 20-28 mm in crown-rump length could be detected through the intact uterine wall even if they occurred. In later fetal life, when higher motor centers with a greater oxygen requirement reach a functional state of development, it may well be that the lower oxygen supply at that time (5) is inadequate for any considerable degree of activity of these higher areas in utero, a possibility that has been suggested by Barcroft and Barron (7).

E. RELATION OF THE DIFFERENT TYPES OF REFLEX ARCS TO THE DEVELOPMENT OF FETAL ACTIVITY

The local limb responses, which are believed by Windle (150, 151) and his colleagues (154, p. 101, for example) to be reflex in nature, are reported in cat fetuses at ages as young as those for which reflexes initiated by trigeminal nerve stimulation are evoked, if not earlier (161, see also footnote, p. 32). As a result these investigators believe that the local limb reflex forms the basis upon which mammalian behavior is developed.

However, contralateral axial flexion in the cervical region, following excitation of the maxillary or mandibular division of V, has been considered by many observers (85) of mammalian fetal behavior to be the mammalian equivalent of the total pattern type of response out of which Coghill found local reflexes to be individuated in Amblystoma. Although accepting the views of Coghill concerning the development of behavior in amphibians, certain investigators of mammalian behavior believe activity in these higher vertebrates to be built up by the integration of local reflexes and have maintained that the generalized responses to trigeminal stimulation are abnormal reactions resulting from anoxia and asphyxia (151). It has been the contention of these observers, then, that only the local limb reflexes elicited under the most favorable conditions (see p. 22) are normal fetal reactions.

The preceding discussion has pointed out that if either of these primary fetal reflexes occurs at all after the onset of anoxia it is normal or characteristic for its type (proprioceptive stretch reflex or exteroceptive flexion reflex) unless decrease in amplitude and synaptic delay render a reflex abnormal (see p. 12). Furthermore, such alterations involve the local limb reflexes following slight amounts of oxygen deprivation even as they do the more generalized responses to trigeminal stimulation after oxygen want increases (150, see also p. 17). Since both the total pattern response of Coghill and the isolated or local reflex of Windle are evidently rendered excitable by an appropriate (although different) degree of oxygen want, and both eventually succumb to the effects of an appropriate (but unlike) degree of asphyxia, to consider one abnormal and the other normal is at least inconsistent.

The misconception that the more generalized trigeminal responses are abnormal in character is undoubtedly related, in part, to the sequence of reactions observed in older fetuses as anoxia progresses. It was stated by Angulo (3) that the fetus recapitulates, in reverse order, the steps in its behavioral development as anoxia develops, that is, the reflexes which have appeared most recently in embryonic development are the first to disappear. Angulo's explanation for this sequence is that the most recently developed reflex centers are the most susceptible to oxygen deprivation and therefore reactions involving such newly developed areas disappear first.

Although reports that the medulla oblongata of newborn mammals has a higher metabolic rate than more recently developed regions of the central nervous system (see 75, pp. 163-164) appear to nullify Angulo's explanation for the sequence of reflex disappearance under anoxia, careful consideration of these studies indicates otherwise. First of all it should be remembered that these accounts are not only based on *in vitro* studies, but that they indicate the metabolic rate of all of the nervous tissue in any region, whether mature or immature, and not the metabolic rate of the functional neurons alone. Such experiments also fail to take into consideration the relative numbers of cell bodies and fibers in an area although cell bodies have a higher metabolic rate than do fiber bundles (57, p. 985). Consequently a low metabolic rate may exist for the higher fetal nervous system centers as a whole, and still be quite in accord with the fact that the functioning neurons of these same regions may have a much higher rate of oxygen consumption. As a result the functional neurons could be suppressed earlier by anoxial conditions than those at lower levels even though the latter may appear to have a higher rate of metabolism demonstrated by these investigations.

In fact other observations indicate that newborn or young mammals actually do have an oxygen gradient for recovery after periods of anoxia (revival time of Gerard; 57, p. 986) which parallels that of the adult. Indeed Windle and Becker (155) found that newborn guinea pigs developed ataxia (cerebellar deficit) and evidences of decerebration after periods of anoxia although retaining normal activity involving brain stem centers. Likewise Nielsen and Courville (122) found cortical or basal ganglion damage in birth asphyxia with the maintenance of brain stem reflex activity. In a more recent paper (153, p. 333) the greater susceptibility of the thalamus, cerebral cortex, and other higher integrative centers of the nervous system of the fetal guinea pig has again been demonstrated by Windle. Inasmuch as the maintenance of central nervous system function under anoxia (survival time of Gerard, 57) parallels the time that nervous system centers can be exposed to anoxia and regain functional status in the adult, a comparable sequence for loss of function under progressing anoxia is to be expected for the fetus.

For the most part Windle and his collaborators have been in agreement with Angulo's explanation (3) for the order of reflex disappearance under the influence of anoxia. However, Windle (151, p. 221 and earlier) has pointed out that the local limb reflexes, which he considers to be the first to appear in the development of mammalian behavior, are also the earliest to disappear under the influence of anoxia (instead of the latest) and hence do not follow this rule. Barcroft and Barron (7, p. 130; 9, p. 436; 10, p. 12) also placed a slightly different interpretation on the changes in mammalian reactions following asphyxia. According to these observers, the movements which reappear as asphyxia progresses do so because of the suppression of the newly developed inhibitory functions of higher brain stem centers which, during adequate oxygenation, have suppressed earlier types of activity.

Obviously certain aspects of these interpretations are in agreement and explain satisfactorily the sequence of reactions as anoxia develops. When examined in the light of the oxygen gradient of the central nervous system and the selective effect of oxygen deprivation on the primary reflex arcs (see pp. 12 and 14), at least the major differences in the various viewpoints may be explained. On the one hand, if the local limb reactions are stretch reflexes which succumb relatively early on exposure to anoxia, and there are many indications that they may be (see p. 17), then, regardless of the age of the fetus or whether or not the reflex first appears early or late in embryonic life, this type of primary reflex will be the earlier of the two primary reflex types to disappear with the onset of anoxia. On the other hand, plurisynaptic exteroceptive reflexes such as reactions to trigeminal nerve stimulation are more resistant to anoxia and will be retained for a longer time. In considering the effect of anoxia on primary fetal reflexes, then, it is the nature of the afferent fibers concerned and not the time of the first appearance of the reflex which determines the reflex survival time under anoxia.

In considering the reactions of older fetuses, in which secondary reflexes have appeared (see p. 11), it should be remembered that brain stem integrative centers, with an oxygen requirement higher than that of all lower motor neurons and primary sensory neurons, participate in the reflexes observed. If the fetal nervous system shows an oxygen gradient which parallels that for the adult, as appears to be the case, then more recently developed higher central nervous system centers will certainly have a higher oxygen gradient and secondary reflexes requiring the participation of such higher centers will be the first to be facilitated, then delayed, and finally suppressed by anoxia before the primary reflex arcs dependent upon the same type of stimuli (i.e., exteroceptive or proprioceptive) are similarly affected.

The assumption that the different types of nerve fibers in the young fetus react to anoxia in the same manner as do their adult counterparts might be disputed on the basis that the neurons are immature and no myelin sheaths are present on any of the fibers at this time. However, even at eight weeks of embryonic age both the cell bodies and the processes of the primary motor and sensory neurons have already acquired certain of their definitive structural characteristics. For example, cell bodies of motor neurons are multipolar and those of the visceral and somatic types are clearly differentiated as to size. Likewise, although only part of the sensory ganglion cells are already changing to the unipolar state, differences in the size of the cell bodies are already present (90, p. 7; 91, p. 342, and Figure 7C). In addition, a difference in size of afferent fibers is present at eight weeks, for the spinal nerve fibers for pain (Lissauer's tract) already appear finer in caliber and have a different staining reaction than do the fibers of the dorsal funiculus (93, p. 40). Inasmuch as it has been indicated that sensory neuronal specificity is determined in part by the peripheral relations of the sensory nerve fibers (129, pp. 236-238; 130, p. 254; 131, pp. 418-420; 140, 143, 144, 149) it may be that the functional characteristics of afferent neurons are decided at the time that the peripheral fibers arrive at their area of distribution. For the proprioceptive neurons this would be when these fibers penetrate the embryonic muscle mass (in 10-12 mm human embryos; 157, p. 496). Certainly the observations of Barron (13) demonstrate that the undifferentiated neuroblasts of the ventral horn begin to develop their morphologic characteristics (as indicated by increase in size and the first appearance of dendrites) after making contact with the myoblasts peripherally. At least the investigators of neuronal specificity have suggested that "lower functional mechanisms" are "qualitatively determined during the developmental . . . phase" (142, p. 92) and not after the onset of function. In view of such factors it is not at all unlikely that a specific fiber type of the young fetus reacts to anoxia essentially in the manner characteristic for adult life even when it begins to function.

Undoubtedly the reappearance of certain fetal reflexes under the influence of anoxia represents a release of the lower motor neurons from the inhibitory effects of higher brain stem centers as was suggested by Barcroft and Barron (7, 10). In other instances, there is probably suppression both of the higher brain stem centers functioning for the local response itself and of other inhibitory centers suppressing the earlier contralateral body flexion response as well, whereas the spinal cord motor neurons with a lower oxygen requirement may be facilitated by the same degree of oxygen lack. The reports of Hooker (83, p. 29) for fetuses of 13½ weeks of menstrual age illustrate this point. Thus Hooker states that although mouth closure may follow stimulation of the lips at this age "fetuses suffering from quite marked oxygen want" exhibit the "more generalized type of neck and trunk responses hitherto observed." So, if only the reflexes dependent upon exteroceptive types of stimuli are considered, these reflexes appear to follow the law of recapitulation in reverse, as formulated by Angulo (3).

The effects of anoxia on proprioceptive reflexes during intermediate stages of fetal development are not reported in enough detail to determine with certainty whether the reverse recapitulation law of Angulo also applies to them as a group. Although Barcroft and Barron (7) were able to elicit tendon reflexes (phasic stretch reflexes) in sheep fetuses at the 60th day of gestation, Hooker (85) has been unable to obtain tendon reflexes in the human fetus until it can be resuscitated, at which time a sufficiently high oxygen supply can be maintained. However, Hooker (83, p. 26) does report proprioceptive responses to stretching muscles at 9½ weeks and Minkowski

(121) describes reactions to muscle stretch at 12 weeks which he considers equivalent to tendon reflexes. The various observations on mammalian fetal behavior do not indicate whether or not the proprioceptive stretch reflexes can be evoked after secondary proprioceptive reflexes (see p. 11) such as righting reflexes have succumbed to asphyxia. The writer has also been unable to discover whether or not the stretch reflexes are retained after asphyxia has blocked righting reflexes in the adult. However, the work of Barcroft and Barron (7, p. 118; 9, p. 441) does indicate that sustained movements, which are suggestive of maintenance of tonus through higher centers, revert to the earlier jerky type, under asphyxia, before disappearing. Such observations at least suggest that the proprioceptive reflexes of the fetus may also follow Angulo's law of recapitulation in reverse. In other words, the proprioceptively elicited reflexes are probably abolished in reverse order following one sequence whereas the exteroceptively evoked reflexes conform to Angulo's reverse law in a different sequence. If this were true, then the labyrinthine reflexes which appear later, according to Barcroft and Barron (8) and Windle (149, p. 178; 158, pp. 177-178), than do the righting reflexes dependent upon changes in head-neck position should be suppressed by asphyxia before the neck righting reflexes are affected and they in turn before the stretch reflexes themselves are lost.

It should be noted that although Lloyd (107, p. 313) states that the "two-neuron-arc pathways are reserved for mediation of the myotatic reflex" (see also 108, p. 324) and many others support this concept (for example, 125, 136) Barron and Matthews (14) suggested earlier that all of the comparatively early discharges may be monosynaptic. Moreover, Brooks and Fuortes (25, p. 94) have pointed out only recently that the assumption of Lloyd is "valid only in limited experimental conditions" and that "the evidence cannot be considered final as yet." Brooks and Fuortes also call attention to the possibility that the "early group of discharges evoked by cutaneous nerve stimulation" may be either "late monosynaptic discharge or an interneurone-mediated reflex."

Moreover, although monosynaptic neuron arcs do transmit stretch reflexes it does not necessarily follow that all stretch reflexes are always of monosynaptic nature. Indeed two recent investigations (77, 115) dealing with normal primates (77, macaque; 115, man) either unanesthetized (77) or with light enough anesthesia to leave the internuncials functional (115), suggest the possibility that internuncials may participate in the stretch reflexes. Thus Magladery and his colleagues (115, p. 132) have found that certain stretch reflexes (designated as *F* reflexes) behave somewhat differently than do the typical monosynaptic reflexes in that the most rapid of "the afferent fibers subserving the response were somewhat slower," the threshold higher, and the reflex itself more resistant to the effects of ischemia. These authors also point out that a mild but ambulatory tabetic with absent ankle jerks and no detectable monosynaptic responses may retain these *F* reflexes (115, p. 146 et seq.). It may be relevant also that Hines and Knowlton (77, pp. 100-101) found "two or more action potentials" with different latencies for each patellar tendon reflex in young macaques during the first three or four years of life but only the potential with the shortest latency in the adult. Although the reflex arc for stretch reflexes in the young fetus has been illustrated here as a purely monosynaptic reflex (Figure 1, broken lines), it may be that stretch reflexes during developmental stages are mediated by plurisynaptic reflex arcs as well.

Spontaneous movements of fetuses have been described frequently, and the term has been applied in various ways. In this discussion the usage of Barcroft and Barron (7), Carmichael (31), and Hooker (85), by which the term is limited to movements of reflex nature for which the stimulus is unknown, is followed. It seems possible that some of the so-called spontaneous movements may be the result of stretch stimuli resulting from the effects of pressure, such as the weight of the embryo pressing upon the surface upon which it rests, as suggested by Windle (158, p. 169). In other instances (30, p. 118; 32, p. 127) the stretch stimulus may be provided by the body bending produced by the limitations of the amnion itself and so lead to the spontaneous extension seen on release of the embryo from the amnion. In any event these spontaneous movements have been observed more often in cat embryos, with the fetus in amnio and the placental circulation intact, than in human embryos where the placental circulation cannot be maintained (83). If certain of the spontaneous movements are stretch reflexes with their low threshold for facilitation on oxygen deprivation, then it seems likely that the small amount of decrease in oxygenation of the blood incident to opening and manipulating the uterus may lead to facilitation and, in consequence, to the appearance of such movements. Conversely, failure to observe spontaneous movements more often in the studies of human fetal behavior (83, 85) may be due to the fact that these observations have been made after placental separation and removal of the fetus from the amnion, that is, at a time when stretch reflexes would be less likely to be elicited. It is interesting that Windle (149, p. 166) has suggested that anoxemia partially accounts for the appearance of spontaneous movements.

Although the controversial points concerning fetal behavior have centered about the development of activity in mammals (12) the appearance of proprioceptive reflexes in submammalian forms should also receive consideration here. That the initial basic reaction of the embryonic nervous system in Amblystoma is a generalized trunk response involving all neuromuscular structures capable of reacting was demonstrated by Coghill (38, 39). This reflex is in response to exteroceptive stimulation (a hair lightly touching the skin) and consists of contralateral trunk flexion limited to the cervical region. The early axial flexure of this type increases in extent in older specimens and is replaced still later by an even more extensive contralateral flexure resulting in the formation of a *C*-shaped coil by the embryo. If no other stimulus follows, the larva often remains in this coil, apparently until the neuromuscular mechanism fatigues. The formation of such a coil must result in a marked tension on the axial musculature on the convex side of the embryo, thus providing a stretch stimulus which should be more than adequate to activate the proprioceptors. However, at the coil stage of behavioral development there is no response to this stretch stimulus. This lack of response is interpreted by the writer to mean that only the exteroceptive fibers of the sensory system are functional at this time. Later the embryo no longer responds by forming a coil but instead goes into the *S*-reaction.

In his earlier papers Coghill (34, p. 100; 35, p. 201; 74, p. 82) expressed the belief that this change from the coil to the *S*-reaction is dependent upon the development of the proprioceptive system. In his later papers Coghill (37, p. 131; 39, p. 991; see also 73, p. 238) favored development of the motor mechanisms alone as the cause of this alteration in activity in Amblystoma. This later view is more in accord with the experiments of Weiss (141, p. 593) in which the spinal cord of tadpoles was de-afferented throughout but the animals still remained able "to swim as aptly and coordinately as normal animals." Evidently the swimming is 'normal' only temporarily and up to a certain point (see also 142, pp. 77-80), however, for in the next paragraph Weiss adds that "chronically de-afferented limbs by and by develop a certain disturbance" which he refers to as "some sort of tonic rigidity . . . which interferes with the normal coordination." Although such experiments demonstrate that "*locomotor coordination can develop by central self-differentiation of the motor centers,*" as Weiss makes clear elsewhere (142, p. 79) they fail to show that sensory impulses exert no influence on the development of locomotion under normal circumstances. In fact, although Weiss' experiments clearly indicate that it is a normal coordination pattern which develops when the larval amphibian is de-afferented, they point out equally well that the movements participating in this coordination pattern become demonstrably abnormal in character in that they soon develop such disturbances as tonic rigidity. It may be that both motor mechanisms and stretch afferents normally participate in the development from the coil stage of Amblystoma to the *S*-reaction, but it is at least clear that if the stretch reflex mechanisms were functional at the time of coil formation, such a coil could neither be formed nor maintained by the embryo.

The presynaptic sensory neuron active at the time of coil formation in Amblystoma is the intramedullary sensory ganglion cell of Rohon-Beard which is often generalized in function, since the peripheral process may divide to send one branch to muscle and one to skin (35, 74). Nevertheless, it is only stimulation of the exteroceptive branch of the peripheral fiber of this sensory neuron which is able to activate muscular contraction in the early flexure and coil stages. Even when the proprioceptive fibers become able to activate the motor neurons, by the very nature of the early sensory apparatus itself—where cutaneous and muscle sensation may be transmitted by the same neuron—responses must be both integrated and generalized in type, in other words, a "total pattern" response or reaction of the body as a whole which has the character of an "avoiding form of reaction" (2, p. 402; 36, p. 283; 93, p. 51). Phylogenetically and ontogenetically, as has been noted elsewhere (12, 90, p. 40), specific, localized reactions can make their appearance only when specificity is developed in sensory receptors and different types of sensory ganglion cells appear. Coghill (35) demonstrated that this is true in the development of Amblystoma for he found that specific reflexes became segregated out of the generalized response at the time when the spinal ganglion cells develop, that is, when different types of sensory neurons, serving different functions, begin to participate.

The proprioceptive functions of the nervous system become of increasingly greater importance during the course of phylogenetic development according to Barron (12) who states that two factors in particular appear to be of paramount importance in the acquisition of greater prominence by this system. One is the change in habitat from water to land and the other is the development of limbs. Both are changes which influence posture and necessitate greater use of the extensor or antigravity muscles. The increased importance of these proprioceptive mechanisms and their consequent early embryonic development is seen even within amphibians. In some frogs (138, Rana guentheri; 163, Rana pipiens, Rana areolata) the larvae may not pass through the coil stage at all, but go directly from the flexure stage to the *S*-reaction. In others (Rana catesbeiana, 163) the coil stage may be identical with that in Amblystoma and in a third group an intermediate step, consisting of a

U-reaction (163), appears at this time. These observations not only support the view that the development of the *S*-reaction is related to proprioceptive impulses, but they also indicate that the proprioceptive system responds to stretch stimuli relatively earlier in some anuran amphibians than in others. As a result, in the course of normal development of certain frogs, formation of a complete coil is prevented (163) apparently because the stretch stimuli cause an early contraction of the axial muscles on the convex side of the embryo before the coil is fully formed.

In this relation it is significant that muscle spindles first make their appearance phylogenetically in such amphibians as the frog and that true nuclei of the dorsal funiculus (nucleus gracilis and nucleus cuneatus) are found first in reptiles (96, pp. 35 and 200). Both of these anatomic advances in the differentiation of the proprioceptive system are in harmony with the increased functional demands on the antigravity muscles imposed on such vertebrates by life on land. Obviously, as the limbs become increasingly important in locomotion and appear progressively earlier in embryonic life, the proprioceptive limb reflexes will likewise develop earlier, a relationship emphasized both by Carmichael (32) and by Barron (12). It is not impossible that localized limb reflexes in response to proprioceptive stimulation may be elicited as early in fetal life as the generalized reaction to exteroceptive trigeminal stimulation, as appears to be true for the guinea pig (31) and as Windle and his collaborators believe (see p. 23). That local reflexes should appear still earlier is unlikely, however, and as yet the number of observations in favor of such an early development of local limb reflexes (160, 161) is too limited to be clearly significant.[5]

Although local limb reflexes are seen as isolated reactions, to consider that they are not integrated in character, either when they first appear, as Windle believes, or at any time so long as the nervous system remains normal, would be to deny the nervous system its most basic function, indeed its primary *raison d'être*. The true nature of the local reflexes as complex, integrated forms of activity has been indicated at various times by both Coghill (40, 41, 42) and Hooker (80, p. 597; 83, p. 7, and 84) and more recently by Sperry (130, p. 237). Many years ago Sherrington (127, p. 145) called attention to the integrative character of the reflex arc itself when he pointed out that each structure at which there may be either "an opportunity for coalition or interference" with neuronal action "is a *mechanism for co-ordination*." Elicitation of the so-called isolated limb response of Windle probably involves at least an inhibition of the activities of the antagonistic muscles and presumably also the axial musculature at the spinal cord levels involved if this reaction is a reflex. Therefore such a reflex, or any other, is integrated into the functional pattern of the nervous system when it first appears, regardless of whether it is manifested early or late in the course of the development of fetal activity.

---

[5] Although it is difficult to determine the exact number of cat embryos for which Windle and his collaborators have observed local limb reflexes at this early period, the number appears to be limited. According to the 1933 paper (160, p. 528) three early embryos which showed local limb reflexes gave no response to trigeminal stimulation, but one which reacted to trigeminal stimulation gave no local limb reflexes. Other specific information has not been found for cat embryos in the papers of these observers. Although emphasizing the need for an "adequate" stimulus in elicitation of limb reactions (150, p. 251) it should be noted that, for the most part, Windle and his colleagues have not used stroking of the skin with a brush or hair in attempting to elicit trigeminal responses, but have employed instead pressure stimuli such as tapping the amnion or direct pressure, types of stimuli usually not effective for the excitation of the exteroceptive nerve fibers which function in the trigeminal reflexes. Consequently one cannot say with certainty that the embryos which showed only local limb reflexes might not also have given trigeminal responses had effective types of stimulation been used.

Even if proprioceptive limb reflexes could be demonstrated satisfactorily as early as Windle believes it would not alter the fact that the generalized response of the entire organism to exteroceptive stimulation is the more primitive type of reaction in mammalian as well as submammalian vertebrates. On the contrary, such precocious development of limb activity in mammals would simply emphasize the progressively increasing importance of the proprioceptive system to higher vertebrates for survival of the organism as a whole. Since the primitive avoiding reaction of contralateral axial flexion (see p. 36) must be replaced, in mammals living on land, by the postural mechanisms represented in the tonic and phasic stretch reflexes, if the newborn animal is to survive, a relatively early development of the function of the antigravity muscles is essential. It should be remembered, however, that this avoiding reaction (or response of the organism as a whole) is merely supplanted and obscured, but not eliminated, as other activities of the nervous system develop. Indeed it is never completely lost, but is retained and even reappears as a widely represented cortical reaction (93). Both the numerous cortical areas from which contralateral (and ipsilateral) body flexion can be elicited in primates (55) and its early developmental appearance on cortical stimulation (76) attest to the primitive nature of this avoiding response. In comparison with this widespread representation of axial flexion in the cortex, independent contraction of individual muscles is restricted to the most highly specialized motor area (area 4).

F. SUMMARY AND CONCLUSIONS

The different types of primary reflex arcs, the two-neuron or monosynaptic reflex arcs related particularly to stretch reflexes and the three- (or more) neuron plurisynaptic reflex arcs which are concerned preeminently with the so-called flexion reflexes, are discussed. A distinction is made between these fundamental reflexes which involve primary motor and primary sensory neurons (and internuncials where present) and other reflexes which secondarily include higher brain stem or even telencephalic centers. It is suggested that the reactions dependent only upon the primary motor and sensory neurons (and the internuncials) be termed primary reflexes, whereas those which also include the integrating mechanisms of the brain stem (such as the reticular formation, for example) or of still higher nervous system centers, be referred to as secondary reflexes, since they appear secondarily in development after the primary reflexes are established. Such a usage would also avoid confusion in the interpretation of fetal behavior studies.

The effects of anoxia and asphyxia upon the primary reflex arcs of the adult are reviewed and the greater resistance of the motor neurons as compared with the presynaptic fibers of the sensory ganglion cells is pointed out. The high degree of susceptibility of the stretch or myotatic reflexes to anoxia is discussed. Attention is called to the fact that the susceptibility of these myotatic reflexes is primarily dependent upon the high oxygen requirement of the large proprioceptive sensory fibers involved. The greater resistance to anoxia of the so-called flexor reflexes which depend upon exteroceptive sensory fibers, all with greater degrees of resistance to anoxia, is likewise presented.

It is pointed out that oxygen deprivation acts on both types of reflexes in the adult to facilitate them initially, then inhibit and finally suppress them as anoxia progresses. The fact is stressed that, in anoxia or asphyxia, a local stretch reflex retains its characteristics in tracings so long as it can be elicited; it varies only as to amplitude and latency of response, and does not become abnormal in any way or take on the peculiarities of the generalized plurisynaptic exteroceptive reflexes. Attention is called to the oxygen gradient of the central nervous system itself, with the high oxygen requirement of the cerebral cortex, cerebellum, and other higher motor centers as compared with the low oxygen need of the lower motor neurons. It is suggested that the primary reflex arcs themselves have a similar oxygen gradient with the most highly specialized sensory fibers most sensitive to oxygen deficit and the lower motor neurons most resistant.

The early mammalian fetal reflexes which have been described are compared with the primary types of adult reflexes. It is concluded that the local reactions of Swenson and Windle may possibly be fetal stretch or myotatic reflexes since they are elicited by stretching a muscle (flipping a limb bud) or pressure (such as produced by tapping the amniotic sac, for example). If they are stretch reflexes these local limb responses are dependent upon proprioceptive afferent fibers with a high oxygen requirement and are mediated particularly by two-neuron or monosynaptic reflex arcs (see p. 17). It is also concluded that the more generalized responses dependent upon stimulation of cutaneous areas supplied by maxillo-mandibular V belong to the so-called flexion reflexes which have an intercalated neuron in the reflex arc and are dependent upon the primitive types of sensory nerve fibers with the lowest oxygen need of all of the sensory fibers.

The relation of anoxia to the interpretation of the various studies on mammalian fetal behavior is discussed. It is pointed out that unless the exteroceptively and proprioceptively elicited fetal reflexes react to oxygen deprivation in a manner totally different from that seen for these reflexes in the adult, both types of reflexes are first facilitated, then depressed and finally blocked entirely as anoxia progresses. Neither one nor the other reflex type will become abnormal. Indeed, except for the difference in the level of reaction to oxygen deprivation, the two types of reflexes evidently respond in the same way, not only to anoxia but also to asphyxia and apparently to some anesthetics.

The misconception of the development of abnormal reflexes in fetuses as the result of oxygen deprivation is discussed. It is suggested that the misunderstanding which has arisen may be the result of the confusion of the primary reflexes, in which the lower motor neurons are the only motor nerve cells involved, with the more complex reactions or secondary reflexes which involve additional and higher integrative motor areas. After secondary reflexes have been added, as higher motor centers become functional, anoxia suppresses these secondary reflexes first, thus producing behavior which may be considered abnormal for the age period of the older fetus, but the primary reflexes which can still be elicited, and which constitute normal activity for a younger fetus, evidently retain their typical characteristics so long as they can be evoked. In studies of fetal activity it is important to remember that until the age at which secondary motor centers have become active, in addition to the primary motor neurons, any true reflexes which can be elicited will remain characteristic under anoxia (or asphyxia) and probably anesthesia.

The development of exteroceptive and proprioceptive reflexes in certain of the submammalian vertebrates is discussed. It is noted that in the coil stage of Amblystoma (38, 39) proprioceptive stretch stimuli, obviously of considerable magnitude, fail to give a reflex response although exteroceptive stimulation of the skin may result in contraction of the same musculature. Since this response is present at a time when only the generalized sensory neurons of Rohon-Beard (which may mediate both cutaneous and muscle sensibility) function as the afferent limb of the reflex arc, it is concluded that the exteroceptive reflexes are ontogenetically older in nature than the stretch reflexes. This primacy of the exteroceptive system was indicated by Sherrington (127, p. 320) when he stated that stimulation of the proprioceptors themselves occurs through action of the organism itself, brought about initially by exteroceptive stimuli.

It is emphasized that all normal nervous system activity is integrated in character. This is true whether the action is a so-called local or isolated reflex of Swenson and Windle or the contralateral axial flexion of Coghill and Hooker. The basic function of the nervous system as an integrator of activity is stressed.

Attention is called to the fact that the stretch reflexes become increasingly important and appear progressively earlier embryologically with the development of limbs and life on land as suggested earlier by Barron (10, 12). It is pointed out that, although this increasingly early development of limbs in higher vertebrates, and hence of stretch reflexes, may result in their appearance almost as early as the exteroceptive responses or even coincident with them, nevertheless the exteroceptively elicited generalized reflex represents, in mammals as well as in submammalian vertebrates, the more primitive type of embryonic behavioral reaction. In other words, the contralateral axial flexure on maxillo-mandibular stimulation in the human embryo is an avoiding reaction of the entire organism (2, p. 402; 36, p. 283; 93, p. 51) which is the mammalian equivalent of the total pattern response first described by Coghill for Amblystoma.

REFERENCES

1. ANGULO (Y GONZÁLEZ), A. W. Endogenous stimulation of albino rat fetuses. *Proc. Soc. Exper. Biol. & Med.*, 1930, **27**, 579.
2. ———. The prenatal development of behavior in the albino rat. *J. Comp. Neurol.*, 1932, **55**, 395-442.
3. ———. Endogenous stimulation of albino rat fetuses. *Anat. Rec., Suppl.*, 1933, **55**, 3-4 (abstr.).
4. ASAYAMA, C. The proprioceptive reflex of a flexor muscle. *Quart. J. Exper. Physiol.*, 1916, **9**, 265-279.
5. BARCROFT, J. The brain and its environment. The Terry Lectures. New Haven, Conn.: Yale Univ. Press, 1938. Pp. 117.
6. BARCROFT, J., & BARRON, D. H. Movements in midfoetal life in the sheep embryo. *J. Physiol.*, 1937, **91**, 329-351.

7. ———. Movement in the mammalian foetus. *Ergebn. d. Physiol., biol. Chemie u. exper. Pharmakol.,* 1939, **42**, 107-152.

8. ———. The development of behavior in foetal sheep. *J. Comp. Neurol.,* 1939, **70**, 477-502.

9. ———. Observations on the functional development of the foetal brain. *J. Comp. Neurol.,* 1942, **77**, 431-454.

10. Barron, D. H. The functional development of some mammalian neuromuscular mechanisms. *Biol. Rev.,* 1941, **16**, 1-33.

11. ———. The early development of the sensory and internuncial cells in the spinal cord of the sheep. *J. Comp. Neurol.,* 1944, **81**, 193-225.

12. ———. Genetic neurology and the behavior problem. In *Genetic Neurology,* ed. by P. Weiss. Chicago, Ill.: Univ. Chicago Press, 1950. (Pp. 223-231.)

13. ———. The differentiation of motor neurons. In *Biological Aspects of Mental Health and Disease.* New York: Hoeber, 1952. (Pp. 186-189.)

14. Barron, D. H., & Matthews, B. H. C. The interpretation of potential changes in the spinal cord. *J. Physiol.,* 1938, **92**, 276-321.

15. Best, C. H.; & Taylor, N. B. The Physiological Basis of Medical Practice, (5th ed.). Baltimore, Md.: Williams & Wilkins, 1950. Pp. 1330.

16. Blair, H. A. Conduction and synaptic transmission in the nervous system. *Ann. Rev. Physiol.,* 1950, **12**, 399-420.

17. Boell, E. J., & Nicholas, J. S. Respiratory metabolism of mammalian eggs and embryos. *Science,* 1939, **90**, 411.

18. Bok, S. T. Das Rückenmark. In *Handbuch der mikroskopische Anatomie des Menschen,* (ed. by G. von Mollendorf). Berlin: Springer, 1928. (Nervensystem, Bd. 4/1, 478-502.)

19. Brooks, C. McC., & Eccles, J. C. Electrical investigation of the monosynaptic pathway through the spinal cord. *J. Neurophysiol.,* 1947, **10**, 251-273.

20. ———. A study of the effects of anaesthesia and asphyxia on the monosynaptic pathway through the spinal cord. *J. Neurophysiol.,* 1947, **10**, 349-360.

21. ———. An electrical hypothesis of central inhibition. *Nature,* 1947, **159**, 760-764.

22. ———. Inhibitory action on a motor nucleus and focal potentials generated therein. *J. Neurophysiol.,* 1948, **11**, 401-416.

23. ———. Inhibition of antidromic responses of motoneurones. *J. Neurophysiol.,* 1948, **11**, 431-444.

24. Brooks, C. McC., Eccles, J. C., & Malcolm, J. L. Synaptic potentials of inhibited motoneurones. *J. Neurophysiol.,* 1948, **11**, 417-430.

25. Brooks, C. McC., & Fuortes, M. G. F. Electrical correlates of the spinal flexor reflex. *Brain,* 1952, **75**, 91-95.

26. Cajal, S. R. Sur l'origine et les ramifications des fibres nerveuses de la moelle embryonnaire. *Anat. Anz.,* 1890, **5**, 85-95.

27. ———. Histologie du système nerveux de l'homme et des vertébrés. Paris: Maloine, 1909. T. 1. Pp. 896.

28. Cannon, W. B., & Burket, I. R. The endurance of anemia by nerve cells in the myenteric plexus. *Amer. J. Physiol.,* 1913, **32**, 347-357.

29. Carmichael, L. The development of behavior in vertebrates experimentally removed from the influence of external stimulation. *Psychol. Rev.,* 1926, **33**, 51-58.

30. ———. Origin and prenatal growth of behavior. In *A Handbook of Child Psychology,* (2nd ed. rev., ed. by C. Murchison). Worcester, Mass.: Clark Univ. Press; London: Oxford Univ. Press, 1933. (Pp. 31-159.)

31. ———. An experimental study in the prenatal guinea-pig of the origin and development of reflexes and patterns of behavior in relation to the stimulation of specific receptor areas during the period of active fetal life. *Genet. Psychol. Monog.*, 1934, **16**, 337-489.

32. ———. The onset and early development of behavior. In *Manual of Child Psychology*, (ed. by L. Carmichael). New York: Wiley, 1946. (Pp. 43-166.)

33. Carmichael, L., & Smith, M. F. Quantified pressure stimulation and the specificity of response in fetal life. *J. Genet. Psychol.*, 1939, **54**, 425-434.

34. Coghill, G. E. The reaction to tactile stimuli and the development of the swimming movement in embryos of Diemyctylus torosus, Eschscholtz. *J. Comp. Neurol.*, 1909, **19**, 83-105.

35. ———. Correlated anatomical and physiological studies of the growth of the nervous system of Amphibia: I. The afferent system of the trunk of Amblystoma. *J. Comp. Neurol.*, 1914, **24**, 161-233.

36. ———. Correlated anatomical and physiological studies of the growth of the nervous system of Amphibia: II. The afferent system of the head of Amblystoma. *J. Comp. Neurol.*, 1916, **26**, 247-340.

37. ———. Correlated anatomical and physiological studies of the growth of the nervous system of Amphibia: VI. The mechanism of integration in Amblystoma punctatum. *J. Comp. Neurol.*, 1926, **41**, 95-152.

38. ———. Anatomy and the Problem of Behavior. New York: Macmillan; Cambridge, England: Univ. Press, 1929. Pp. 113.

39. ———. The early development of behavior in Amblystoma and in man. *Arch. Neurol. & Psychiat.*, 1929, **21**, 989-1009.

40. ———. Individuation versus integration in the development of behavior. *J. Gen. Psychol.*, 1930, **3**, 431-435.

41. ———. Integration and motivation of behavior as problems of growth. *J. Genet. Psychol.*, 1936, **48**, 3-19.

42. ———. Early embryonic somatic movements in birds and in mammals other than man. *Monog. Soc. Res. Child Devel.*, 1940, **5**, 1-48.

43. Coghill, G. E., & Angulo (y González), A. W. The development of the sensory system in relation to the local sign. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1934, **15**, 265-273.

44. Cohen, L. A. Localization of stretch reflex. *Fed. Proc.*, 1952, **11**, 25-26 (abstr.).

45. Corbin, K. B., & Harrison, F. Function of the mesencephalic root of fifth cranial nerve. *J. Neurophysiol.*, 1940, **3**, 423-435.

46. Coronios, J. D. Development of behavior in the fetal cat. *Genet. Psychol. Monog.*, 1933, **14**, 287-386.

47. Creed, R. S., Denny-Brown, D., Eccles, J. C., Liddell, E. G. T., & Sherrington, C. S. Reflex activity of the spinal cord. Oxford: Clarendon Press, 1932. Pp. 183.

48. Cushny, A. R. Pharmacology and therapeutics, (13th ed. rev. by A. Grollman & D. Slaughter). Philadelphia: Lea & Febiger, 1947. Pp. 868.

49. Denny-Brown, D. E., & Liddell, E. G. T. Observations on the motor twitch and on reflex inhibition of the tendon-jerk of M. supraspinatus. *J. Physiol.*, 1927, **63**, 70-80.

50. Eccles, J. C. Synaptic potentials of motoneurones. *J. Neurophysiol.*, 1946, **9**, 87-120.

51. Eccles, J. C., & McIntyre, A. K. Plasticity of mammalian monosynaptic reflexes *Nature*, 1951, **167**, 466-468.

52. Ekbom, K. A., Jernelius, B., & Kugelberg, E. Perioral reflexes. *Neurology*, 1952, **2**, 103-111.

53. Fazekas, J. F., Alexander, F. A. D., & Himwich, H. E. Tolerance of the newborn to anoxia. *Amer. J. Physiol.*, 1941, **134**, 281-287.

54. Fitzgerald, J. E., & Windle, W. F. Some observations on early human fetal movements. *J. Comp. Neurol.*, 1942, **76**, 159-167.

55. Fulton, J. F. Physiology of the nervous system (3rd ed.). New York: Oxford Univ. Press. Pp. 667.

56. Gasser, H. S. Pain-producing impulses in peripheral nerves. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1943, **23**, 44-62.

57. Gerard, R. W. Anoxia and neural metabolism. *Arch. Neurol. & Psychiat.*, 1938, **40**, 985-996.

58. Gesell, R., Hansen, E. T., & Siskel, J. On the electrotonic nature of stimulation, inhibition, summation and after-discharge of nerve centers. *Amer. J. Physiol.*, 1947, **148**, 515-529.

59. Glass, H. G., Snyder, F. F., & Webster, E. The rate of decline in resistance to anoxia of rabbits, dogs, and guinea pigs from the onset of viability to adult life. *Amer. J. Physiol.*, 1944, **140**, 609-615.

60. Glazer, W. Regulation of respiration: XXX. The effects of mechanical asphyxia and administration of carbon dioxide, sodium carbonate, sodium bicarbonate and sodium cyanide on the reflex response of the anterior tibial muscle of the dog. *Amer. J. Physiol.*, 1929, **88**, 562-569.

61. Granit, R. Reflex self-regulation of muscle contraction and autogenetic inhibition. *J. Neurophysiol.*, 1950, **13**, 350-372.

62. Grenell, R. G. Central nervous system resistance; effects of temporary arrest of cerebral circulation for periods of 2 to 10 minutes. *J. Neuropath. & Exper. Neurol.*, 1946, **5**, 131-154.

63. Groat, R. A., & Koenig, H. Centrifugal functional deterioration of asphyxiated motor nerve within the neural axis. *J. Neurophysiol.*, 1946, **9**, 463-469.

64. Grollman, A. Pharmacology and therapeutics. Philadelphia: Lea & Febiger, 1951. Pp. 828.

65. Grundfest, H. The properties of mammalian B fibers. *Amer. J. Physiol.*, 1939, **127**, 252-262.

66. Hagbarth, K.-E., & Naess, K. Reflex effects of tetanic stimulation of different afferent fibre-systems in the hind limb of the cat. *Acta physiol. Scandinav.*, 1950, **21**, 336-361.

67. Van Harreveld, A. On spinal shock. *Amer. J. Physiol.*, 1940, **129**, 515-523.

68. ———. Survival of reflex contraction and inhibition during cord asphyxiation. *Amer. J. Physiol.*, 1944, **141**, 97-101.

69. Van Harreveld, A., & Tyler, D. B. The influence of temperature on spinal cord damage caused by asphyxia. *Amer. J. Physiol.*, 1944, **142**, 32-39.

70. Harrison, F., & Corbin, K. B. The central pathway for the jaw-jerk. *Amer. J. Physiol.*, 1941, **135**, 439-445.

71. Harrison, R. G. An experimental study of the relation of the nervous system to the developing musculature in the embryo of the frog. *Amer. J. Anat.*, 1904, **3**, 197-220.

72. Hawkins, J., McLaughlin, C. R., & Daniel, P. Neuronal damage from temporary cardiac arrest. *Lancet*, 1946, **1**, 488-492.

73. Herrick, C. J. George Ellett Coghill, naturalist and philosopher. Chicago, Ill.: Univ. Chicago Press, 1949. Pp. 280.

74. Herrick, C. J., & Coghill, G. E. The development of reflex mechanisms in Amblystoma. *J. Comp. Neurol.*, 1915, **25**, 65-85.

75. Himwich, H. E. Brain metabolism and cerebral disorders. Baltimore, Md.: Williams & Wilkins, 1951. Pp. 451.

76. Hines, M., & Boynton, E. P. The maturation of excitability in the precentral gyrus of the young monkey (Macaca mulatta). *Carnegie Instit. Wash., Contrib. Embryol.*, 1940, **28**, No. 518, 309-451.

77. Hines, M., & Knowlton, G. C. Patterns of organization of tendon reflexes suggested by knee jerk latencies. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1952, **30**, 98-117.

78. Hoffmann, P. Untersuchungen über die Eigenreflexe (Sehnenreflexe) menschlicher Muskeln. Berlin: Springer, 1922. Pp. 116.

79. ————. Die physiologischen Eigenschaften der Eigenreflexe. *Ergebn. d. Physiol., biol. Chemie u. exper. Pharmakol.*, 1934, **36**, 15-108.

80. Hooker, D. Early fetal activity in mammals. *Yale J. Biol. & Med.*, 1936, **8**, 579-602.

81. ————. Fetal behavior. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1939, **19**, 237-243.

82. ————. Fetal reflexes and instinctual processes. Part of a symposium on *The biology and psychology on "instinctual" processes. Psychosomat. Med.*, 1942, **4**, 199-205.

83. ————. The origin of overt behavior. Ann Arbor, Mich.: Univ. Mich. Press, 1944. Pp. 38.

84. ————. Neural growth and the development of behavior. In *Genetic Neurology*, (ed. by P. Weiss). Chicago, Ill.: Univ. Chicago Press, 1950. (Pp. 212-213.)

85. ————. The prenatal origin of behavior. *Porter Lectures*, Ser. XVIII, March 12-13, 1951. Lawrence: Univ. Kansas Press, 1952. Pp. 143.

86. ————. Human fetal activity. In *Biological Aspects of Mental Health and Disease*. New York: Hoeber, 1952. (Pp. 189-192.)

87. Hooker, D., & Nicholas, J. S. Spinal cord section in rat fetuses. *J. Comp. Neurol.*, 1930, **50**, 413-467.

88. Houssay, B. A. Human Physiology. New York: McGraw-Hill, 1951. Pp. 1118.

89. Howell, W. H. A Textbook of Physiology. (16th ed. rev. by J. F. Fulton). Philadelphia: Saunders, 1949. Pp. 1258.

90. Humphrey, T. Primitive neurons in the embryonic human central nervous system. *J. Comp. Neurol.*, 1944, **81**, 1-45.

91. ————. Intramedullary sensory ganglion cells in the roof plate area of the embryonic human spinal cord. *J. Comp. Neurol.*, 1950, **92**, 333-400.

92. ————. The caudal extent of the descending root of the trigeminal nerve during the period of early human fetal activity (8 to 8.5 weeks of menstrual age). *Anat. Rec., Suppl.*, 1951, **109**, 46-47 (abstr.).

93. ————. The spinal tract of the trigeminal nerve in human embryos between 7½ and 8½ weeks of menstrual age and its relation to early fetal behavior. *J. Comp. Neurol.*, 1952, **97**, 1-67.

94. Johnston, J. B. The radix mesencephalica trigemini. *J. Comp. Neurol.*, 1909, **19**, 593-644.

95. Jolly, W. A. On the time relations of the knee-jerk and simple reflexes. *Quart. J. Exper. Physiol.*, 1911, **4**, 67-87.

96. Kappers, C. U. Ariëns, Huber, G. C., & Crosby, E. C. The Comparative Anatomy of the Nervous System of Vertebrates, Including Man. New York: Macmillan, 1936. Pp. 1845.

97. Kato, G. The Microphysiology of Nerve. (New ed.). Nihonbashi, Tokyo, Japan: Maruzen, 1934. Pp. 139.

98. Kaya, R., & Starling, E. H. Note on asphyxia in the spinal animal. *J. Physiol.*, 1909, **39**, 346-353.

99. King, C. E., Garrey, W. E., & Bryan, W. R. The effect of carbon dioxide, hyperventilation, and anoxemia on the knee jerk. *Amer. J. Physiol.*, 1932, **102**, 305-318.

100. Von Koelliker, A. Handbuch der Gewebelehre des Menschen. (6th ed.) Leipzig: Engelmann, 1896. Bd. 2, 874.

101. Laporte, Y., & Lloyd, D. P. C. Disynaptic reflex linkage between muscles of a myotatic unit. *Fed. Proc.*, 1951, **10**, 78-79 (abstr.).

102. Von Lenhossék, M. Der feinere Bau des Nervensystems im Lichte neuester Forschungen. Berlin: Kornfeld, 1895. Pp. 409.

103. Liddell, E. G. T., & Sherrington, C. S. Reflexes in response to stretch (myotatic reflexes). *Proc. Roy. Soc., London*, 1924, **96B**, 212-242.

104. ———. Further observations on myotatic reflexes. *Proc. Roy. Soc., London*, 1925, **97B**, 267-283.

105. Lloyd, D. P. C. A direct central inhibitory action of dromically conducted impulses. *J. Neurophysiol.*, 1941, **4**, 184-190.

106. ———. Reflex action in relation to pattern and peripheral source of afferent stimulation. *J. Neurophysiol.*, 1943, **6**, 111-119.

107. ———. Neuron patterns controlling transmission of ipsilateral hind limb reflexes in cat. *J. Neurophysiol.*, 1943, **6**, 293-315.

108. ———. Conduction and synaptic transmission of reflex response to stretch in spinal cats. *J. Neurophysiol.*, 1943, **6**, 317-326.

109. ———. Functional organization of the spinal cord. *Physiol. Rev.*, 1944, **24**, 1-17.

110. ———. Facilitation and inhibition of spinal motoneurons. *J. Neurophysiol.*, 1946, **9**, 421-438.

111. ———. Integrative pattern of excitation and inhibition in two-neuron reflex arcs. *J. Neurophysiol.*, 1946, **9**, 439-444.

112. ———. Establishment and relief of asphyxial blocks in spinal motoneurons. *Fed. Proc.*, 1950, **9**, 80-81 (abstr.).

113. ———. On reflex actions of muscular origin. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1952, **30**, 48-67.

114. Lorente De Nó, R. Analysis of the activity of the chains of internuncial neurons. *J. Neurophysiol.*, 1938, **1**, 207-244.

115. Magladery, J. W., Parks, A. M., Porter, W. E., & Teasdall, R. D. Spinal reflex patterns in man. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1952, **30**, 118-151.

116. Magoun, H. W., & Rhines, R. Spasticity, the Stretch-Reflex, and Extrapyramidal Systems. Springfield, Ill.: Thomas, 1947. Pp. 59.

117. Matthews, B. H. C. Nerve endings in mammalian muscle. *J. Physiol.*, 1933, **78**, 1-53.

118. Matthews, S. A., & Detwiler, S. R. The reactions of *Amblystoma* embryos following prolonged treatment with chloretone. *J. Exper. Zool.*, 1926, **45**, 279-292.

119. McCouch, G. P., Deering, I. D., & Stewart, W. B. Inhibition of knee jerk from tendon spindles of crureus. *J. Neurophysiol.*, 1950, **13**, 343-350.

120. McDonald, R. A. Anoxia: its effect on the physiology and biochemistry of the brain and on behavior. In *Biological Aspects of Mental Health and Disease*. New York: Hoeber, 1952. (Pp. 335-357.)

121. Minkowski, M. Neurobiologischen Studien am menschlichen Foetus. *Abderhaldens Hdbh. d. biol. Arbeitsmeth.*, 1928, Abt. V, T. 5B, H. 5, 511-618.

122. Nielsen, J. M., & Courville, C. B. Role of birth injury and asphyxia in idiopathic epilepsy. *Neurology*, 1951, **1**, 48-52.

123. Porter, E. L., Blair, R. K., & Bohmfalk, S. W. Facilitation of reflex action in the spinal cat following asphyxia of the cord from cyanide and increased intra-spinal pressure. *J. Neurophysiol.*, 1938, **1**, 166-175.

124. Quastel, J. H. Effects of drugs on metabolism and physiologic activity of brain. In *Biological Aspects of Mental Health and Disease*. New York: Hoeber, 1952. (Pp. 360-376.)

125. Renshaw, B. Activity in the simplest spinal reflex pathways. *J. Neurophysiol.*, 1940, **3**, 373-387.

126. ———. Observations on interaction of nerve impulses in the gray matter and on the nature of central inhibition. *Amer. J. Physiol.*, 1946, **146**, 443-448.

127. Sherrington, C. S. The Integrative Action of the Nervous System. New Haven, Conn.: Yale Univ. Press; London: Oxford Univ. Press, 1906. Pp. 411.

128. ———. Notes on the scratch-reflex of the cat. *Quart. J. Exper. Physiol.*, 1910, **3**, 213-220.

129. Sperry, R. W. Neuronal specificity. In *Genetic Neurology*, (ed. by P. Weiss). Chicago, Ill.: Univ. Chicago Press, 1950. (Pp. 232-239.)

130. ———. Mechanisms of neural maturation. In *Handbook of Experimental Psychology*, (ed. by S. S. Stevens). New York: Wiley; London: Chapman & Hall, 1951. Pp. 1436.

131. Sperry, R. W., & Miner, N. M. Formation within sensory nucleus V of synaptic associations mediating cutaneous localization. *J. Comp. Neurol.*, 1949, **90**, 403-423.

132. Sprague, J. M., Schreiner, L. H., Lindsley, D. B., & Magoun, H. W. Reticulo-spinal influences on stretch reflexes. *J. Neurophysiol.*, 1948, **11**, 501-507.

133. Strassmann, P. Das Leben vor der Geburt. *Samml. Klin. Vortr.*, N. F., Nr. 353, (Gynak. Nr. 132), 1903, 947-968.

134. Strong, O. S., & Elwyn, A. Human Neuroanatomy. (2nd ed.) Baltimore, Md.: Williams & Wilkins, 1948. Pp. 442.

135. Swenson, E. A. The development of movement of the albino rat before birth. Doctorate Thesis, 1926. Dept. Anat., Univ. Kansas, Lawrence. (Unpublished.)

136. Szentagothai, J. Anatomical considerations of monosynaptic reflex arcs. *J. Neurophysiol.*, 1948, **11**, 445-454.

137. Thelander, H. E. The course and distribution of the radix mesencephalica trigemini in the cat. *J. Comp. Neurol.*, 1924, **37**, 207-220.

138. Wang, G.-H., & Lu, T.-W. Development of swimming and righting reflexes in frog (Rana guentheri): Effects thereon of transection of central nervous system before hatching. *J. Neurophysiol.*, 1941, **4**, 137-146.

139. Weinberg, E. The mesencephalic root of the fifth nerve. A comparative anatomical study. *J. Comp. Neurol.*, 1928, **46**, 249-405.

140. Weiss, P. Selectively controlled central-peripheral relations in the nervous system. *Biol. Rev.*, 1936, **11**, 494-531.

141. ———. Does sensory control play a constructive role in the development of motor coordination? *Schweiz. med. Wochenschr.*, 1941, 71st year, No. 12, 591-595.

142. ———. Self-differentiation of the basic patterns of coordination. *Comp. Psychol. Monog.*, 1941, **17**, 1-96.

143. ———. Lid-closure reflex from eyes transplanted to atypical locations in Triturus torosus. Evidence of a peripheral origin of sensory specificity. *J. Comp. Neurol.*, 1942, **77**, 131-169.

144. ———. An introduction to genetic neurology. In *Genetic Neurology*, (ed. by P. Weiss). Chicago, Ill.: Univ. Chicago Press, 1950. (Pp. 1-39.)

145. ———. Central versus peripheral factors in the development of coordination. *Res. Publ. Ass. Nerv. & Ment. Dis.*, 1952, **30**, 3-23.

146. Wiggers, C. J. Physiology in Health and Disease. Philadelphia: Lea & Febiger, 1949. Pp. 1242.

147. Windle, W. F. Correlation between the development of local reflexes and reflex arcs in the spinal cord of cat embryos. *J. Comp. Neurol.*, 1934, **59**, 487-505.

148. ———. On the nature of the first forelimb movements of mammalian embryos. *Proc. Soc. Exper. Biol. & Med.* 1937, **36**, 640-642.

149. ———. Physiology of the Fetus: Origin and Extent of Function in Prenatal Life. Philadelphia: Saunders, 1940. Pp. 249.

150. ———. Genesis of somatic motor function in mammalian embryos: a synthesizing article. *Physiol. Zool.*, 1944, **17**, 247-260.

151. ———. Reflexes of mammalian embryos and fetuses. In *Genetic Neurology*, (ed. by P. Weiss). Chicago, Ill.: Univ. Chicago Press, 1950. (Pp. 214-222.)

152. ———. Asphyxia Neonatorum: Its Relation to the Fetal Blood, Circulation and Respiration and Its Effects Upon the Brain. Springfield, Ill.: Thomas, 1950. Pp. 70.

153. ———. Anoxia: its effect on structure of the brain. In *Biological Aspects of Mental Health and Disease*. New York: Hoeber, 1952. (Pp. 327-334.)

154. Windle, W. F., & Becker, R. F. Relation of anoxemia to early activity in the fetal nervous system. *Arch. Neurol. & Psychiat.*, 1940, **43**, 90-101.

155. ———. Effects of anoxia at birth on central nervous system of the guinea pig. *Proc. Soc. Exper. Biol., & Med.*, 1942, **51**, 213-215.

156. Windle, W. F., Becker, R. F., Rhines, R., & Cowgill, E. J. Effect of anoxia on trigeminal reflexes in cat fetuses. *Physiol. Zool.*, 1942, **15**, 375-382.

157. Windle, W. F., & Fitzgerald, J. E. Development of the spinal reflex mechanism in human embryos. *J. Comp. Neurol.*, 1937, **67**, 493-509.

158. Windle, W. F., & Griffin, A. M. Observations on embryonic and fetal movements of the cat. *J. Comp. Neurol.*, 1931, **52**, 149-188.

159. Windle, W. F., Minear, W. L., Austin, M. F., & Orr, D. W. The origin and early development of somatic behavior in the albino rat. *Physiol. Zool.*, 1935, **8**, 156-185.

160. Windle, W. F., O'Donnell, J. E., & Glasshagle, E. E. The early spontaneous and reflex behavior in cat embryos and fetuses. *Physiol. Zool.*, 1933, **6**, 521-541.

161. Windle, W. F., Orr, D. W., & Minear, W. L. The origin and development of reflexes in the cat during the third fetal week. *Physiol. Zool.*, 1934, **7**, 600-617.

162. Yanase, J. Beiträge zu Physiologie der peristaltischen Bewegungen des embryonalen Darmes: II. Beobachtungen an menschlichen Föten. *Arch. f. d. ges. Physiol.*, 1907, **119**, 451-464.

163. Youngstrom, K. A. Studies on the developing behavior of Anura. *J. Comp. Neurol.*, 1938, **68**, 351-379.

Department of Anatomy
School of Medicine
University of Pittsburgh
Pittsburgh 13, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 45-58.

THE PREPARATION OF CASE BALANCE SHEETS*

Department of Psychology, New School for Social Research

KENNETH IVES

A. INTRODUCTION

In order to tell what the effects of an experience have been on a person, a more comprehensive and objective means for describing the status of the person before and after it is needed. This is basic both to the evaluation of treatment programs and to the analysis of the results of social forces and social movements upon the people they affect.

Probably the simplest and most inclusive measure of effects through the analysis of case records, the "movement scale" (5) admits to certain serious limitations (6, pp. 42-43). It fails to provide a measure of status, by which different cases could be matched, and it does not distinguish the factors which have changed, although it is based upon an analysis of them. An earlier measure, the Distress-Relief Quotient (DRQ) (2) provides the former factor in a limited way, but not the latter. However, it correlates poorly (.2) with the movement scale (4, p. 81).

Personality inventories, batteries of tests, and projective tests are also used as a means of assessing status and changes in a case. These are uneven and incomplete in their coverage. The factors they report on are not standardized in terminology and content between the various tests. Thus four different factors are referred to as "introversion-extroversion" (3).

In an effort to develop a more adequate and objective means of describing a case, the writer has used as a model methods and concepts adapted from accounting, with which he is familiar. In economic activities, accounting has long provided an inclusive and detailed description and measurement of the single case and its component factors.

The most basic accounting report is variously called the "balance sheet," the "position statement," or the "statement of assets and liabilities." It shows the resources of, and demands on, the unit under study at the time stated. As adapted to psychological analysis, assets and liabilities are factors which, respectively, aid or hinder the expansion of the personality. This dimension appears to include Lewin's concept of "life space" (10, pp. 42-52),

---

*Received in the Editorial Office on August 1, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

Jennings' concept of "social space" (9, p. 219), and the concept of "movement" referred to above.

B. SCOPE

This paper describes and illustrates a method of analysis and tentative categories applied to the preparation of balance sheets on a published, summarized case record. While it is believed that the framework here described can provide a means for the analysis and coordination of factors and results of various tests and inventories, this approach has not been developed much as yet, and will not be described here. If more extended description is desired of accounting, its adaptation to the present material, and the sources from which the categories were developed, these may be found elsewhere (7, pp. 5-27, 67-90).

C. THE SAMPLE CASE

The case chosen for illustrative purposes is one reported by Combs (1). It gives a summarized record of the six interviews with the subject. This meets the criterion of minimum length for the analysis of movement, as found in case work research (6, pp. 23ff.). There are about 350 items (phrases, sentences, incidents) in the summarized case record.

Bernreuter Personality Inventory scores were obtained shortly before and about a year after treatment. The changes in these scores are striking. Major changes in the subject's behavior and outlook are also reported in the closing interview. The large changes described, the presentation of test and other data supplementing the evidence in the case record, and the brief yet adequate length, all make this a case well adapted for intensive study.

D. METHOD

The method of analysis builds in part on a previous use of content analysis on case records (2, pp. 22-28). It consists of the steps listed in Table 1.

TABLE 1
STEPS IN PREPARING CASE BALANCE SHEETS

1. Establishing the points of reference: person and times.
2. Dividing the case record into units (items).
3. Classifying each item as a balance sheet, movement or other item.
4. Numbering the items, for reference.
5. Selecting the account into which each balance sheet item fits.
6. Copying each balance sheet item into its account.
7. Correcting the accounts, by "journal entries".
8. Summarizing the items in the accounts, by "journal entries".
9. Translating them from the language of the source into an objective language of relationships.
10. Transferring these relationships to summary accounts.
11. Preparing statements from the summary accounts.

For illustration of the material and the application of the steps to it, a page early in the sample case is given in Table 2. For reasons of space and clarity only the major steps will be discussed here.

The most important single step is fitting the items in the case to the categories (accounts), the fifth step. Definitions of the categories are given in Table 3. They were developed from factors common to several fields.

TABLE 2
SAMPLE PAGE OF CASE RECORD

	Class	Acct.	#	*
C. You've been pretty much under the domination of your parents./	x			
T. When I'm home I'm under their power./ If I	o	SL	II 20	*
went back home I'd fall right back into it./ When I	o	EL	21	*
didn't go home last summer /	o	SD+	22	*
I was all right./ As soon as I got there everything was	o	RL	23	*
back the same way again./	o	DL	24	*
C. You feel that last summer you gained somewhat?/	x			
T. I guess it's one of the disadvantages of/ being	m		25	
almost the youngest child. / I think they're trying to	o	SL	26	*
keep hold of us. / I don't know—maybe I'm just the	o	EL	27	*
black sheep of the family!	m		28	
C. You feel that you're the black sheep. /	x			
T. Um, hum. Why even when I was chosen queen	m		29	
of Maryland county / they would not let me stay in	o	SL	30	*
the county seat at Kent. / My parents came right along	o	SL	31	*
with me. / I had the feeling they thought I'd run away. /	o	SR	32	*
They're always right there. / They didn't want me to	o	SL	33	*
go to school either. / My mother writes and says, "I'll	o	SR	34	*
certainly be glad when you're out of Midwestern University." /	o	SR	35	*
C. These things upset you pretty much. /	x			
T. Yes, they do. / The bull sessions with the other	o	DG	36	*
girls—it just doesn't seem to be the same. / They can	o	PO	37	*
tell all about their parents, / and tell their parents	o	PO	38	*
things,/	o	PO	39	*
but not me—/	o	PS	40	*
I have to keep things bottled up. /	o	BI	41	*
I don't even dare tell my sister. /	o	PS	42	*
C. It's pretty necessary to keep things bottled up. /	x			
T. I've just started to let loose. / All these things	o	BE	43	*
may seem trivial and silly to worry about, / but they	o	ER	44	*
do bother me awfully. / Others wonder why I can't tell	o	DG	45	*
my parents anything. /	o	DB	46	*
But they would think it was bad. / My folks probably	o	ER	47	*
know I'm not telling them things. / Maybe that's the	o	PO	48	*
reason for their holding on to me so. / Maybe if I did	m		49	
tell them things more they'd loosen up. (pause) / You'd	m		50	
think in three years I'd lose this feeling. / Doesn't it	m		51	
sound as if I should? / Why should this overpowering	m		52	
hold of dad's still be on me? /	m		53	

TABLE 3

DEFINITIONS OF CATEGORIES

Social Environment

The Social Environment accounts record behavior and emotional reactions towards the person by others, as he is in contact with them.

Acceptance Received, (SA); (asset). Acceptance ranges from passive toleration to active affection. It includes both recognition and response, and all associative processes directed towards the person.

Rejection Received, (SR); (liability). All dissociative processes and attitudes, such as distrust and dislike, directed towards the person and towards groups, activities and other expressions with which he feels emotional identification.

Submission of Others, Higher Status, (SH); (asset). Acceptance by others of leadership by the person. This includes their willingness to do so even when he fails to provide the leadership.

Domination by Others, Lower Status, (SL); (liability). All efforts by others to achieve control over the person, whether successful or not; includes punishment and destructiveness. Efforts by others to get the person to run his life by their advice or directives.

Distance, (SD+ and SD—). Distance here is not primarily geographic, but social and psychological. It is measured by: frequency and length of contact; the extent of communication, sharing and influence while contact is made; and the emotional content of the contacts. Distance from a source of rejection or domination is an asset; from a source of acceptance, from a person chosen or desired, is a liability. Approachability of a person in the environment is an asset, their busyness or avoidance of the subject is a liability.

Social Behavior Patterns

These accounts refer to the actions of the person towards others.

Association, (BA); (asset). Habits of contacting and working with others: cooperativeness, social identification and sharing.

Dissociation, (BD); (liability). Habits of avoidance and isolation from others. Negative valence expressed towards others, their ideas and activities. Not to be confused with conflict, here thought to be a mixed process of associating and dominating tendencies.

Extroversion, (BE); (asset). The habit of meeting the environment with initiative, spontaneity, efforts at domination. The expression of one's own feelings, thoughts and urges. Acceptance and exploration of the social environment. Includes both communicativeness and activity; the direction outwards of drives and responses. Habit of leading in both association and dissociation. Behavior is placed in this account on the basis of its immediate effects only.

Introversion, (BI); (liability). The habit of meeting the environment with inhibition and repression of expressive urges; acquiescence, submission, avoidance of communication; the direction of drives inwards. Note that this may result in either learning, fantasy or self-suppression, with widely different long-range results. Nevertheless, it is first placed in this liability account.

Adaptive Abilities

These accounts refer to evidences of the various skills useful for effective functioning in any area of social life. Thus they are all considered to be assets.

Perception of Others, (PO). Observation and interpretation of their activities and motives, especially towards the subject. Understanding their experiences and behavior, especially as compared to one's own.

Perception of Self, (PS). Awareness of one's own actions, emotions, motives, abilities, possibilities and limitations. Attitudes to self.

Perception of Causal Relationships, (PC). Insight, understanding, functional or structural vision. Sensing of the interrelations between experiences, feelings, motives and activities, on the basis of which future behavior can be predicted accurately.

TABLE 3 (*continued*)

Organization, (AO). Wishes, plans, initiative in planning for self, responsibility accepted for own actions and condition. Includes the social activities a person relates to biological urges or undertakes for their fulfillment. The positive drive *towards,* separated from the other drives in Dollard and Mowrer's distress group. Includes "should" and "ought" responses. These usually indicate external relationships which the person's internal organization feels urged to develop.

Methods for the Control of Others, (AC). Technics of influence, leadership, education and training used by the person for achieving their wishes and for meeting and redirecting the influences of the social environment.

Social Skills, (AS). All other aptitudes and experiences which the person can use in meeting various social situations.

Social Expectations

These accounts refer to what the person expects to receive from the social environment in the future, both in general and in response to specific behavior on his part. They include the social attitudes which determine his plans for future situations. The first of each pair is an asset, the second a liability.

Acceptance vs. *Rejection Expected,* (EA & ER). Expectations of "belonging" and being at ease, vs. of being ostracised, attacked, threatened or ridiculed.

Submission of Others vs. *Domination by Others Expected,* (EH & EL). Higher vs. lower status expected in the future. Superiority and inferiority complexes. Obligations felt from and to others. Rights and duties felt. Faith in a "magic helper" or "an improvement in luck" vs. fears or delusions of persecution.

Expectations of Success or *Failure* of own efforts, (ES & EF). Confidence, feelings of strength or weakness in various areas.

Relief—Distress

These accounts record the emotional reactions within the person resulting from the factors shown in the other groups of accounts. Relief items are assets, distress ones are liabilities.

Affection vs. *Hostility,* (RA & DH). Positive or negative valence of person's feelings towards other people, activities and groups.

Generalized Relief or *Distress,* (RG & DG). Joy, feelings of success, vs. sorrow, misery, frustration, where these do not refer to particular causes.

Relief or *Distress Localized,* (RL & DL). The same, but related to specific factors or causes. These two sets of accounts are internalized reactions, whereas affection and hostility are directed outwards.

Blockages and Resistances, (DB): (liability). Inability to explore or operate in certain areas of living. "I can't" reactions. May result from person's failures or felt weakness, or from prohibitions by others.

Compulsions and Projections, (DP); (liability). The conversion or transference of activity or emotion from one field or several, into another field where it is inappropriate. It is important to distinguish this transference from the change in attention from one aspect of a problem to another relevant and productive aspect or technic for making progress on it. Theories underlying these "conversion" reactions include: that the perceptive abilities of the person are too weak to permit identification of the source of distress or relief; that specific blockages to identification exist; or that the capacity of existing channels of expression are not adequate to express the amount of emotion generated.

Conflicts and Ambivalences, (DC); (liability). These are considered as distress items because they require a considerable input of energy with little or no net output of results, even tho they are sometimes a step in the substitution of a better for a poorer response.

Principal sources, in addition to those already cited from social work research (2, 5), were in psychological testing (12), sociology (14), psychiatry (11), and non-directive therapy (13). This approach was used partly to avoid the limitations, semantic and otherwise, of trying to describe a personality from the point of view of only one field of inquiry or school of thought. It was also used so that the resulting framework could fit into studies in any of the fields, and could help to clarify interrelations between them as shown in particular cases.

E. Primary Analysis

With the aid of the definitions, items can be placed in the appropriate categories. For example, Items II 30 and 31 (Table 2): "they would not let me stay in the county seat at Kent. / My parents came right along with me./" Here "Alice Taylor" is describing the behavior of her parents towards her, so they belong in the "social environment" group of categories. In the first item, domination by her parents is specifically referred to. As this limits her range of activity, it is a liability. It fits one of the definitions of the account with the initials "SL," so these were put in the second column. In the next item the domination is not so evident. The effect of her parents' coming along is to prevent her from acting independently, to maintain her dependent status and their control over her. Hence this item was also marked "SL."

A dozen items later she refers to her own behavior (II 43): "I've just started to let loose." This places it in the "behavior" group. The expression of her own thoughts and feelings which it refers to is an expansion of her personality, is found on the asset side, and was placed in the "extrovert behavior" (BE) account.

Just before this (II 38-39) she tells about the other girls at college: "They can tell all about their parents, / and tell their parents things." As it is not their relations to her which she describes, but their relations with their parents, these items were classed as perception (PO) rather than in the "environment" group. Seeing how others act in a similar situation is a resource she can use, to compare her own behavior with or to imitate. Thus it is an asset, a means for taking steps to expand her ways of acting.

While the use of the future tense is a clear indication that the "expectations" group of accounts is referred to, some items are more obscure. Consider Item II 44: "All these things may seem trivial and silly to worry about. . . ." Here she indicates that she expects rejection of her problems by the use of the devaluative terms "trivial" and "silly." These are not her own valuations of the problems as the next item (II 45) indicates: "but they do bother me awfully." In this item the distress is directed to a variety of problems, so it was placed in the generalized distress (DG) account. Concern over her own problems, and the feeling that she cannot share them freely, are both factors which limit her social expansiveness, and are thus liabilities.

F. MULTIPLE REFERENCES

One of the problems which arises in the use of this procedure is the handling of items which refer to more than one account. Take for example an item in the fifth interview, "I don't dare do anything without their approval." Taken as a whole, this item reflects a habit of dependence on the attitudes of her parents, which she expects will continue. Thus it was first placed in the account for "expectations of lower status" (EL).

However, a closer inspection shows that parts of it refer to four other accounts as well. The fact that she is describing her own attitude indicates perception of self (PS). Not daring to act under certain conditions is evidence of blockage (DB). Relying on her parents' decisions instead of her own shows introvert behavior in the sense used in this paper. And the need for approval of her behavior in advance implies that otherwise she expects she will receive criticism or other forms of rejection (ER) from them.

The most frequent type of multiple reference problem is in items which refer both to evidence of perception by the person, and also to the content of what was perceived.

G. INTERMEDIATE STEPS

After all the items referring to the opening and closing situations have been marked for the categories (accounts) whose definitions they fit, they are copied into those accounts. Separate sets of accounts are used for the opening and closing items. Sample accounts are given in Table 4. Several items already discussed, and others from Table 2 may be found there. At this point the items in each account should be checked, to discover errors in classifying, or multiple references, and the necessary changes made.

The tabulation of items into the accounts provides a thorough basis from which to summarize the factors, using all relevant material. The mechanics of summary are shown in Table 5. Similar items in an account are grouped together under one description, and are transferred to a summary account by means of a "journal entry." This is a balancing device from which the material included is checked off in the original account, and a description of it is recorded in a summary account. If there is much material, two stages of summary may be used, as in this case, or even more. This permits each

TABLE 4
SAMPLE ACCOUNTS

Item	#	Domination by others—lower status (SL)	Debit	Credit
II	3	When I lived at home, I was overseered by my father or mother	JE 17	#
	4	I was told about everything	JE 17	#
	9	If I had a date, dad would know all about it before I got home	JE 18	#
	8	(Every one knew) how strict dad was	JE 18	#
	17	Dad made me (remain) just a child	JE 18	#
	18	Ever since I was a young child I've always been left under his hands	JE 18	#
	19	My parents always did my thinking for me	JE 17	#
	20	When I'm home, I'm under their power	JE 17	#
	26	Being almost the youngest child	JE 19	#
	30	They would not let me stay (alone) at the county seat	JE 17	#
	31	My parents came right along with me	JE 17	#
	33	They're always right there	JE 17	#
III	77	Dad was strict and overpowering all the time	JE 18	#
II	57	Parents are strict in little things as well as big	JE 17	JE 4
		Extroversion (BE)		
II	13	When I was a freshman, I let my hair grow	#	JE 24
	43	I've just started to let loose	#	JE 24
V	36	I told them I wanted to work	JE 1	JE 24
		Perception of Causal Relations (PC)		
II	23	When I didn't go home last summer, I was all right	JE 13	JE 39

stage of summary to be simple and clear, and yet produce a final one-page summary or "balance sheet." Such a clear and concise summary is desirable because the inferences which can be made from any complex data depend upon the number of significant factors which can be observed in operation at one time.

H. SECONDARY ANALYSIS

In connection with the summarizing process, the second major analytic step occurs. It is the translation of the material from the language of the source into an objective language of social relationships. This is done by the basis upon which items are grouped for summary, and the resulting explanation of the summarizing journal entry. The explanation should list the process in operation, the people related by it, and the direction of the process between them. In Journal Entry 18, Table 5, "Domination received from father" does this. "Alice Taylor" being the point of reference does not need to be mentioned. It is this functional explanation which is carried forward to the summary account.

TABLE 5
SUMMARIZING PROCEDURE

SUMMARIZING JOURNAL ENTRIES	*	Debit	Credit
		(from Lower Status account, Table 4)	
17			
Domination by others—Items II 3, 4, 19, 20, 30, 31, 33, 57.	SL	#	
Summary Account—Situation	SS		#
To combine items in which domination is received from both parents.			
18			
Domination by others—Items II 8, 9, 17, 18; III 77.	SL	#	
Summary Account—Situation	SS		#
To combine items in which domination is received from father.			

SUMMARY ACCOUNTS

Social Situation (SS)	#	Db.	Cr.
Rejection received from people in general	2	JE 68	JE 14
Rejection received from parents	10	JE 68	JE 15
Domination received from both parents	8	JE 69	JE 17
Domination received from father	5	JE 69	JE 18
Status of almost youngest child	1	JE 69	JE 19
Parents unapproachable	1	JE 68	JE 22

SUMMARIZING JOURNAL ENTRIES	*	Db.	Cr.
69			
Social Situation,—JE's 17, 18, 19.	SS	#	
General Summary	GS		#
To combine domination received from parents, especially father.			(to General Summary account)

Items can now be traced through from the case record to the balance sheet. Thus Item II 30 referred to before, is shown as given by the subject in Table 2. It was placed in the account "SL" in Table 4. Then, with seven other items from that account, it is summarized by Journal Entry 17, Table 5 and transferred into the "situation summary" account shown there. These and all other items from the "SL" account are further summarized by Journal Entry 69 into one item, "Domination received from parents, especially father" which is carried forward to a general summary account (not shown) and appears in that form on the balance sheet, Table 6.

TABLE 6

OPENING BALANCE SHEET: CASE OF "ALICE TAYLOR"

Assets		Liabilities and Net Worth	
Social Situation			
Few advantages in situation	3	Rejection received from parents and others	13
		Domination received from parents, especially father	14
Social Behavior			
Few efforts at extrovert behavior	3	Inhibition of expression	5
		Introversion in relations with parents	7
Adaptive Abilities			
Perception of others' attitudes on communication	10		
Perception of own shortcomings	9		
Perception of causal relations	3		
Desire for communication and independence	2		
Control of others by tantrums	1		
Social Expectations			
		Rejection expected, especially from father	9
		Domination expected, especially from parents	3
		Feelings of insignificance	6
		Dependent status will continue	8
		Failure expected	3
Relief—Distress			
Relief at distance from parents	1	Hostility to parents	3
		Distress, especially at home situation and vocation	18
		Blockage of communication and initiative	15
		Conflicts, mostly internal	16
Net Worth			
		Bernreuter Personality Inventory; average of all six factors (F factors only, 14%)	8%
		Relief—Distress Quotient	2%
		Scholastic performance	1.03

I. THE BALANCE SHEETS

By a thorough application of the series of steps to the sample case record, the balance sheets at opening and closing of treatment shown in Tables 6 and 7 were obtained. The numbers after each factor show how many items

TABLE 7
CLOSING BALANCE SHEET: CASE OF "ALICE TAYLOR"

Assets		Liabilities and Net Worth	
	Social Situation		
Acceptance received from father	2		
Adult status granted by father	1		
Close to (accepting) parents	4		
	Social Behavior		
Extroversion towards parents, especially father	3		
	Adaptive Abilities		
All opening perceptive abilities and desires	24		
Additional perceptive abilities	3		
Vocational choice and decision	4		
Responsibility for own problems	5		
Desire for extrovert behavior	5		
Control by meeting domination squarely	1		
	Social Expectations		
Expectation of success and of higher status	3		
	Relief—Distress		
Affection for father and home	4	Uncertainty as to cause of the improvement	1
Generalized relief	11		
Relief at change relations with parents	4		
Relief at independence acquired	2		
Relief at sharing with parents	3		
Relief at counseling process	3		
	Net Worth		
		Bernreuter Personality Inventory; average of all six factors (F factors only, 44%)	40%
		Relief-Distress Quotient	96%
		Scholastic performance	1.34

in the original case record were summarized into it. Detailed comparison of these opening and closing balance sheets will show major changes in almost every factor.

The "Net Worth" sections contain whatever indicators can be found of the general ability of the person to function socially. In theory, this section represents the balance between the asset and liability factors. As these are not now measurable in interchangeable units, the balance cannot be obtained directly by subtraction, as is done in business accounting. Outside evidences such as test scores and performance ratings can be used where available. The evidences shown will thus be incomplete and spotty at best.

In this case, the Bernreuter Personality Inventory scores obtained before and a year after treatment were averaged by a method developed by the author (8) and form one evidence of net ability to function. More objective evidence is provided by the scholastic averages of "Alice Taylor" for the two semesters preceding and following counseling (1, p. 154) respectively.

The Relief-Distress Quotient is an internal indicator, being the percentage that relief items are of relief plus distress ones. It is an inversion and partial redefinition (7, pp. 74-75) of the DRQ developed by Dollard and Mowrer (2). Other internal indicators based on the number of items could be prepared and might prove significant.

One should be cautious in drawing conclusions based solely on the relative number of items in various categories, however. For example, the most significant change in this case may be in the reported means of control over others, which is represented by only one item in each balance sheet. In the opening picture, "Alice Taylor" observed, "I used to hold my breath . . . as a way of getting my way." While this refers to childhood, it is possible that she still used some similar but less obvious emotional technique. In the closing interview she reports, "(Father and I) had a hot and heavy argument for quite a while," apparently the first time this had happened. After that incident, her father's attitude and behavior towards her changed. This change in her technique may therefore have been the key to most of the other ones shown by the closing balance sheet.

J. Reliability and Internal Consistency

Reliability of the classification of the items into the categories was tested by an independent judge, and by a retest made by the author over a month after previous inspection of the material. On the major differentiation, that between assets and liabilities, agreement with the independent judge was .85, while the retest figure was .92. The majority of the differences were caused by the perception-content double reference problem referred to earlier. Consistent handling of that would have raised the reliability figures to .96 and .97 respectively.

Internal consistency in the results on this case is the only indication of validity yet available. The improvement reported by the counselor shows as an increase in the adaptive abilities and in the few evidences of "net worth," as well as by a shift in all other groups from the liability to the asset side.

K. ADVANTAGES

Comparison of this method and its resulting case summaries with those used in the article in which the sample case was presented will show major improvements in clarity, definiteness, and inclusiveness.

The condensation of material into one-page summaries; the division of these into six groups, two sides and about 30 factors; and the use of functional definitions for the factors, all make it easier to see the whole case and the relations of the parts. This is important because the theoretical and diagnostic interpretations which can be made from complex data depend on the number of significant factors which can be perceived at one time. Similarly, the bringing together of related items into an account is an aid to perceiving the quality, intensity, and direction of that factor, and its relation to other factors. .

As compared to other more or less graphic analyses, the balance sheets include a wider area. A sociometric chart covers only some aspects of behavior and environment. A personality profile covers some abilities and some behavior patterns. These and other factors are all included in a complete balance sheet. Such an inclusive coverage provides a more adequate basis for tracing all of the factors in a given situation, and discovering hidden or secondary results of the changes observed. The full list of factors can be used as a check upon the completeness and adequacy of the information supplied about a case.

The inclusion of the distress-relief factors is important because these contribute largely to the strength and course of labor, farmer, political and religious movements, as well as to less socially organized behavior of the individuals who feel them. The expectations people have about future events and about the effects of present policies also affects their behavior strongly. It affects the distress-relief ratio through formula that happiness is, in part, results less expectations. These two types of factors are rarely shown and estimated in case summaries.

The steps by which the balance sheets are prepared form a tight fitting and explicit chain from the original material to the final reports. Errors, omissions, and differences in interpretation can be specifically located and corrected for.

The use of balance sheets as a basis for evaluating changes in a case or for comparing different cases should be more thorough and generally much faster than when other types of case reports are used. By placing two balance sheets side by side, individual factors and the general patterns can be compared almost at a glance.

L. SUMMARY AND CONCLUSIONS

The present paper has applied to a published, summarized case record a new framework and method for case analysis and summary. The principal steps are described and illustrated with material from the case record. Status of the case at opening and closing of treatment are given in the two balance sheets prepared. Striking differences in nearly every factor are shown. Reliability of the major differentiation is reported. Advantages of the method and of the balance sheets over case analysis methods and case reports now in use are suggested.

In view of the advantages claimed and illustrated, further development, testing, and use of this method of case analysis should prove fruitful.

REFERENCES

1. COMBS, A. W. Follow-up of a counseling case treated by the non-directive method. *J. Clin. Psychol.,* 1945, **1**, 147-154.
2. DOLLARD, J., & MOWRER, O. A method of measuring tension in written documents. *J. Abn. & Soc. Psychol.,* 1947, **42**, 3-32.
3. GILLILAND, A. R. What do introversion-extroversion tests measure? *J. Abn. & Soc. Psychol.,* 1933-34, **38**, 407-412.
4. HUNT, J. McV. Measuring the effects of social casework. *Trans. N. Y. Acad. Sci.,* 1947, Ser. II, **9**, 78-88.
5. ——— *et al.* Measuring Results in Social Casework: A Manual on Judging Movement. New York: Family Service Asso. of America, 1950.
6. ———, BLENKNER, M., & KOGAN, L. S. Testing Results in Social Case Work: A Field Test of the Movement Scale. New York: Family Service Asso. of America, 1950.
7. IVES, K. H. An introduction to the application of accounting methods to social science data: the preparation of case balance sheets. Unpublished Masters Thesis, New School for Social Research, 1950.
8. ———. A note on the presentation of personality inventory scores. *J. Clin. Psychol.,* 1950, **6**, 415-416.
9. JENNINGS, H. H. Leadership and Isolation. New York: Longmans Green, 1943.
10. LEWIN, K. Principles of Topological Psychology. New York: McGraw Hill, 1936.
11. LASSWELL, H. D. A provisional classification of symbol data. *Psychiatry,* 1938, **1**, 197-204.
12. MASLOW, A. H., *et al.* A clinically derived test for measuring psychological security-insecurity. *J. Gen. Psychol.,* 1945, **33**, 21-41.
13. SNYDER, W. U. An investigation of the nature of non-directive psychotherapy. *J. Gen. Psychol.,* 1945, **33**, 193-223.
14. WIESE, L. VON, & BECKER, H. Systematic Sociology. New York: Wiley, 1932.

314 West 94 Street
New York City 25

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 59-80.

THE ASSESSMENT OF COMMUNICATION: I. A METHOD FOR THE ANALYSIS OF SOCIAL INTERACTION*[1]

Division of Psychiatry of the University of California School of Medicine, and The Langley Porter Clinic, San Francisco

JURGEN RUESCH, JACK BLOCK, AND LILLIAN BENNETT

A. THE PROBLEM

In recent years, the importance of communication has been increasingly recognized. In the fields of cybernetics (9, 10), psychology (4), psychiatry (5), group dynamics (3), and public opinion (7), concern with communication has yielded promising results. In effect, communication theory has emerged as a theory of theories (2) because it is concerned with the scientist's processes of perception and evaluation of nature. While in the past scientists have taken the processes of perception and evaluation for granted and have assumed that a scientific report reflects largely the events in question, communication oriented theorists, ranging in label from psychiatrists to mathematicians and engineers, have called attention to the fact that any body of information is largely dependent upon the human communication system within which it is gathered. Communication theory has united a number of fields, not so much because of similarities in the subject matter but because of similarities in the way information is collected, corrected, evaluated, and applied.

In the narrower sphere of psychology and psychiatry, communication theory has enabled the clinician to reformulate the questions probing into abnormal behavior. Ruesch (6), for example, has called attention to the fact that the vast majority of terms used in psychiatry refer to the communicative behavior of patients, and that, in fact, all psychopathology can be viewed as a disturbance of communication. If this assumption should prove to be productive, then a first step has been taken to observe directly in clinical practice those processes which have both disruptive and therapeutic effects. By focussing upon communication, we describe observable ongoing events rather than end products or end stages of processes. We

---

*Received in the Editorial Office on August 11, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This investigation was supported in part by a research grant from The National Institute of Mental Health of the National Institutes of Health, Public Health Service.

come closer to the relevant data and we free clinical concepts and terminology from some of the mentalistic encumbrances and reifications of the past.

A new viewpoint raises new questions, and new questions require new methods. Human communication can only be examined in a social context—in a situation where at least two persons participate. One person does not function alone when gathering information about people since even observing means interacting in some way with the persons under observation. The first principle underlying human communication theory, therefore, holds that the data of observation or testing have to include the communicative behavior of all the persons participating in the social situation, irrespective of whether they are observers or test subjects. The second principle of communication theory views the exchange of messages as a constantly ongoing process of feedback and correction. Notoriously, variables with a slow rate of change often do not reproduce the dynamic qualities inherent in human relations—processes which are basically feedback phenomena.

In terms of the above two criteria, it was quickly apparent that conventional methodology was inadequate for the task. The majority of psychological tests tacitly treat the individual as a self-contained entity functioning without regard for social context and the personality of the examiner. The observer as an integral part of the method is not usually considered. Furthermore, most of the available approaches emphasize the long lasting and relatively slow changing features of personality.

Our own approach was oriented around the idea that communication is achieved when correspondence of information has been established between two or more people. This, of course, is an ideal condition which in daily life rarely obtains. In most instances, we have to content ourselves with partial correspondence, ranging from complete misunderstanding to rather imperfect understanding. However, the degree of correspondence of information established between two people in the course of communication is a function amenable to controlled observation and measurement. Our procedure for the assessment of communication has focussed upon this variable and its fluctuations as an index relevant for the understanding of social interaction.

The method utilizes statements printed on cards—that is, language—to reflect the impressions which the exchange of messages has made upon people. The statements describe feelings, opinions, attitudes, and social actions. Utilizing Stephenson's Q-sort procedure (8), the participants in a communication system are asked to sort the cards along a bipolar continuum wherein the statements evaluated as most characteristic (typical, frequent.

intense) are placed at the one end and the statements evaluated as least characteristic (atypical, infrequent, weak) are placed at the other. The innovation introduced by this method is seen neither in the type of statements used nor in the principle of sorting but in the fact that each person or each group sorts two sets of cards, the first bearing upon one's own person or group, the second upon the other person or group. By correlating the sortings, data are obtained which express the degree of agreement present. The investigator then is enabled to specify the particular principles underlying the areas of disagreement between people or groups. A variety of statistical and qualitative comparisons of the sortings furnishes information on the ways in which people structure their mutual exchange of messages and organize their interactions.

B. Theoretical Considerations

In making any scientific observation, it is necessary first to define the frame of reference. It is obvious that the universe of discourse may be treated or "sliced" in various ways. We have chosen to consider the problem of communication and interaction in terms of four levels of complexity. This approach is based upon the following theoretical considerations.

Human beings are equipped with instruments of communication which we may conveniently subdivide according to their functions of reception (perception, input-channels), evaluation (cognition, decision-making, and including memory), and transmission (action, output-channels). The individual's capacity in each of these functions is limited. Due to these limitations in the ability to observe and abstract, an individual who focusses upon one person will see this person in greater detail than if he focusses upon 25 people. If he focusses upon a thousand people, he will be unable to keep them separate and the picture will be blurred. The closer an individual comes to an object (or objects), the more detail he will see per object while maintaining his total perceptual capacity constant. If the observer is distant and tries to embrace many objects and many people, the individual characteristics of each item will be lost and all that will be abstracted are relatively gross inferences, usually statistically derived. In brief, the larger the system a person wishes to observe, the more generalized must be the observations. These observations are determined not only by the size of the communication system which is being observed but also by the observer's position within the network.

Foci or levels of communication can be described as follows:

1. *The intrapersonal level.* The focus upon one person includes ob-

servations in terms of traits, attitudes, values, habits, and all the enduring characteristics of the individual.

2. *The interpersonal level.* The focus upon two people includes observations in terms of actions, motivations, intents and effects of messages, and moods occurring in two-person situations.

3. *The intragroup level.* The focus upon a small group includes observations in terms of the internal structure of the group and the organization of communication.

4. *The intergroup level.* The focus upon at least two small groups includes observations in terms of the social functions and effects of groups upon each other.

It should be emphasized that an individual's focus of orientation may change and change abruptly. In a group, for example, an individual may speak at one instant for himself and at another instant as a group member. The foci of orientation, therefore, have to be viewed in a constant process of oscillation, a fact which makes the scientific analysis of these phenomena difficult. Our own approach has been to keep the four levels separate, testing the individual first at one level and then at another. The distortion introduced is one of sequence. In one context, we assume an individual adopts an intrapersonal focus; in another context, we assume he is adopting an intergroup focus, et cetera. While granting that this procedure disregards the subject's actual oscillation of foci, it is to be noted that we may interrelate the various levels and approximately regain this information.

Messages may be destined for home consumption (intero-functions) or for consumption by an outside receiver (extero-functions). For example, an individual may talk to himself (intrapersonal level) or he may direct a message to his wife (interpersonal level). As a club member, he may have a message for his fellow (intragroup level), or on occasion he may represent his group vis-a-vis another group (intergroup level). Depending upon whether the message is intended for an internal destination or for an external destination, the content and form of the message change. In other words, both the sender and the receiver of a communication will formulate the content and form of a message according to the focus they possess at a given moment.

1. *Assumptions Made in the Test Construction*

The structuralization of the test into four levels has been carried through in the following manner: For example, the statement at the intrapersonal level, "I am critical of myself," is matched at the interpersonal level by a statement, "I am critical of him," at the intragroup level with the statement, "We are critical of each other," and at the intergroup level with the statement, "We are critical of them." In carrying these statements through four levels, we have made a series of assumptions.

1. We assume that the verbal statements printed on cards and subsequently sorted by the individual are sufficiently representative of the events which take place at the time of an interaction.

2. We assume that the memory and evaluative capacity of the subject are such as to allow him to abstract and condense his experience into the terms which are presented on the cards. For these reasons, all statements are confined to simple sentences utilizing popular words.

3. We assume that the words, "I am critical of myself," at one level really have a correspondent meaning to the statement "I am critical of him" at another level. Molar equivalence as well as molecular equivalence is required.

4. We assume that the normal distribution forced upon individuals' sortings will approximately fit their natural tendencies of classification. Though it must be granted that there are individuals who tend to classify asymmetrically, the assumption applies for most individuals. For most of the continuistic psychological universes, something like a normal distribution is a realistic assumption and, of course, is statistically convenient. Some sort of fixed distribution is required in order to prevent the loss of information which would result from idiosyncratic, asymmetric distribution of cards; all things considered, the normal distribution seems best.

Support for these assumptions must be in terms of the reliability of the sortings and the validity or coherence of the findings. A completed study offers encouraging evidence (1).

2. *The Nature of the Statements*

The tests, except for the Communication Test, consist of statements that people make about themselves, about others, and about interacting with others. The statements selected are a small but relevant sample from those available. Our selection has been determined by a number of factors, the most important of which is language. The popular words in which interactional statements are phrased range from slang, which concisely tends to "hit the nail on the head," to sophisticated, intellectual, or "psychological" terminology, which often is the only form in which complexities and subtleties of behavior can be expressed. Our selection excludes these extremes as much as possible and is restricted to "plain language" without local cultural coloring and at a level which the average person can understand. The use of "plain language" does not by any means eliminate ambiguities, and although our selection has been secondarily determined by an effort to minimize ambiguity there are included in the tests statements such as "I help him" which can be interpreted by the sorter in a number of ways.

We have used almost exclusively, short sentences referring to single actions, feelings, and attitudes, even though recognizing that communication is frequently characterized by multiple, simultaneous messages If statements such as "He attracts and rejects me at the same time" were included, the number of possible combinations would make the procedure unwieldy However, the analyzer can obtain an idea of such combinations by scanning clusters of statements for their complex interrelationships.

An additional criterion for the selection of statements was the "declinability" of the verbs like "I disappoint," or of adjectives and adverbs used in conjunction with verbs like "I am lonely." The statements had to be adapted to the four levels. Furthermore, the statements included not only the subject, or actor (I, he, we, they), but also the object of the action (I-myself and he-himself; I-him and he-me; we-each other and they-each other; we-them and they-us). The first 63 statements in all the tests use the same key words and vary only with regard to the pronouns. Statements 64-85 are parallel in all tests except the intrapersonal version, while the remainder are not comparable to each other. These latter statements are more or less specific for each level, and if declined would lose meaning. Table 1 will clarify our procedure.

The statements selected refer not only to overt, descriptive "action" such as "I argue with him" and "He belittles me," but also to feelings and attitudes which are conveyed and evoked. Some of these statements describe an action which effects a specific response, whereby the "action" statement implies the effect, such as "I antagonize him." Some action statements, when preceded by "try to," denote the participant's intent and his acknowledgment of responsibility for interpersonal events. Some statements bear upon personality "traits" which are brought out only in certain interpersonal contexts; e.g., "I express myself clearly when with him." Furthermore, some statements have been included which refer to interpersonal activities carried out in fantasy; e.g., "I worry about him."

The collection of statements, then, can be classified into the following categories:

Statements referring to action

1. Simple action statements of varied levels of abstraction. Examples: "I argue with him," "I reproach him."

TABLE 1

The Asssessment of Communication: Schematic Representation

Type of interaction	Intrapersonal		Interpersonal		Intragroup		Intergroup		
Label	"I"	"He"	"I-him"	"He-me"	"We"	"They"	"We-them"	"They-us"	
Code number	A	B	AB	BA	C	D	CD	DC	
Card number 1	I am critical of myself	He is critical of himself	I am critical of him	He is critical of me	We are critical of each other	They are critical of each other	We are critical of them	They are critical of us	example
↓ 63	etc.	etc.	etc.	etc.	etc.	etc.	etc.	etc.	
64	Paired statements characteristic of intrapersonal functioning		I have a lot in common with him	He has a lot in common with me	We have a lot in common with each other	They have a lot in common with each other	We have a lot in common with them	They have a lot in common with us	example
↓ 85			etc.	etc.	etc.	etc.	etc.	etc.	
86 ↓ 100			Paired statements characteristic of interpersonal functioning		Paired statements characteristic of intragroup functioning		Paired statements characteristic of intergroup functioning		

2. Actions which denote intent. Example: "I try to reassure him."
3. Actions which infer intent. Example: "He tries to reassure me."
4. Actions with built-in effect. Example: "He embarrasses me."
5. Action with inferred effect. Example: "I bore him."

Statements referring to feelings

6. Subjective feelings. Example: "I like him."
7. Inferred feelings. Example: "He is ill at ease with me."

Statements referring to attitudes or expectations

8. Subjective attitudes. Example: "I respect him."
9. Inferred attitudes. Example: "He trusts me."

Interactive statements bearing upon personality "traits"

10. Interpersonal trait variability. Example: "I have difficulty making decisions when with him."

C. Description of the Method

The over-all purpose of the method is to gather information about one person as he evaluates and functions in a variety of two-person and group situations. Agreement and disagreement reached with other people are checked by means of mutual sortings. While this problem is covered in our design by the Intrapersonal, Interpersonal, Intragroup, and Intergroup Tests, an additional procedure, the Communication Test, considers the individual and his communicative behavior independently of his social relations. This latter card set collects data relevant for theoretical issues.

While a survey of the whole method is given in Table 1, a word needs to be said about the parallelisms which have been established.

At each level, two tests are paired. Statements referring to the structure of the self ("I") are paired with statements which refer to the structure of the other person ("He"); statements referring to what "I do to him" are paired with statements which analyze what "he does to me." Statements which refer to the internal group structure, or the "We," are paralleled with statements which refer to the "They"; and, finally, what "we do to them" is paired with what "they do to us." The Intrapersonal and Intragroup Tests are relatively structural and refer to attributes of a more permament nature; the Interpersonal and Intergroup Tests are more dynamic and refer to differentiated social action and its effects.

Depending upon the question under investigation and the clinical situation, the researcher will use the four tests selectively. The tests can be used alone (two sets at each level) or in combination.

The first 63 statements in all versions of the test utilize the same words except for the pronouns. This means that comparisons between levels can be carried out by correlating the scores assigned to the first 63 cards at one level with the scores assigned to the first 63 cards at another level. Cards 64-85 utilize the same words in the Interpersonal, Intragroup, and Intergroup Tests so that those 22 statements can also be correlated between levels. The rest of the statements are relatively idiosyncratic and specific for each test and cannot be correlated directly.

1. *The Process of Card Sorting*

The subject is required to sort the cards into nine categories containing 2, 5, 12, 19, 24, 19, 12, 5, and 2 cards respectively. These card frequencies approximate a normal distribution, an arrangement which is both reasonable in its assumptions and convenient for analysis. The basis of card sorting is in terms of a continuum ranging from those cards deemed "most characteristic," "most relevant," "most typical" on the one hand to those cards deemed "least characteristic," "least relevant," "least typical" on the other. It should be noted that a statement placed as "least characteristic or relevant" by a sorter may be as important to him (and to the investigator) as a statement placed as "most typical." An individual's sorting is a function of his evaluation of the intensity, frequency, and importance of the behavioral events described on the cards. The judged importance of an item is determined by the sorter's value system and may supersede frequency and intensity in the determination of relevance. Reliabilities of the order of .8 attest to the meaningfulness of the complex weighting procedure the sorter uses.

In practice, the examiner instructs the subject to start with a dichotomous or trichotomous sorting of the cards into fairly gross categories. Successive dichotomizations of the cards are directed by the examiner so that ultimately the various categorizations may be recombined with the cards ordered into a nine-step continuum. Subjects find successive dichotomization an easier task than dealing with nine steps at one time. The principle of ordinality insures the meaningfulness of the ultimate recombining.

Preferably, both sets of cards for the level under consideration are sorted on the same day by the subject. The "He," the "He-me," the "They," and the "They-us" series are always sorted first in an attempt to sensitize the sorter to the characteristic features of the interaction he is describing and in order to develop a set which can be carried over to his sorting of the "I," the "I-him," the "We," and the "We-them" tests. The subject is suitably oriented and reminded every so often that he is to consider himself specifically with regard to the designated other person or group.

TABLE 2
The Intrapersonal Test: "I" Version*

A- 1 I am critical of myself.
2 I get along well with myself.
3 I like to be by myself.
4 I understand myself.
5 I like myself.
6 I inhibit myself.
7 I belittle myself.
8 I have conflicts with myself.
9 I give in to myself.
10 I encourage myself.
11 I try to outdo myself.
12 I try to reassure myself.
13 I am kind to myself.
14 I discipline myself.
15 I trust myself.
16 I think about myself.
17 I bore myself.
18 I respect myself.
19 I underestimate myself.
20 I am honest with myself.
21 I rely upon myself.
22 I am careful not to upset myself.
23 I know my limitations.
24 I am tolerant of contradictions in myself.
25 I am afraid of myself.
26 I help myself.
27 I have mixed feelings about myself.
28 I protect myself.
29 I have difficulties making decisions.
30 I am interested in myself.
31 I teach myself.
32 I embarrass myself.
33 I control myself.
34 I deceive myself.
35 I argue with myself.
36 I am emotional.
37 I disappoint myself.
38 I overestimate myself.
39 My feelings towards myself change from day to day.
40 I make myself unhappy.
41 I get annoyed at myself.
42 I expect a lot of myself.
43 I get impatient with myself.
44 I try to change myself.
45 I wear myself out.
46 I worry about myself.
47 I am lonely by myself.
48 I am important to myself.
49 I act against my own interests.
50 I am ill at ease with myself.
A- 51 I operate at a fast rate.
52 I entertain myself.
53 I make plans.
54 I am moralistic with myself.
55 I have faith in myself.
56 I function well by myself.
57 I reproach myself.
58 I am at my worst by myself.
59 I surprise myself.
60 I am pleased with myself.
61 I try to influence myself.
62 I know what I want.
63 I push or drive myself.
64 When I am in doubt, I weigh all the alternatives.
65 I have a sense of humor.
66 I am confused.
67 I think about my past.
68 I think about my future.
69 I have many talents.
70 I am interested in gadgets.
71 I have a good memory.
72 I am self-sufficient.
73 There are many sides to my personality.
74 I bounce back after disappointments.
75 No matter how badly things go, I never give up.
76 I'm an optimist.
77 I never let an opportunity go by.
78 I have high ideals.
79 I am a perfectionist.
80 I feel inadequate.
81 I take chances.
82 I can adapt to circumstances.
83 I let my feelings get the best of me.
84 I have no pride.
85 I am a day-dreamer.
86 I am original and imaginative.
87 I am an impulsive person.
88 It bothers me when things upset my daily routine.
89 I am a slave to my conscience.
90 I am a playful person.
91 I take myself seriously.
92 I have a nasty disposition.
93 I'm attractive.
94 I'm a maladjusted person.
95 I am a clear thinker.
96 I function well under stress.
97 I am fidgety and restless.
98 I am ambitious.
99 I am sloppy.
100 I complete the things I start.

---

*For the "He" version, substitute the words "he" for "I" and "himself" for "myself."

2. *The Intrapersonal Test*

The Intrapersonal Test consists of 100 cards (see Table 2) on which are typed statements describing a variety of personality features. The test comes in two versions: the "I" version for self-evaluation and the "He" version for the rating of another specified person. The statements are not interactional in nature, but bear upon features thought to be characteristic of the individual. This test taps the somewhat idealized and generalized aspects of a person's character. The test can be used by itself in order to follow an individual's self-evaluation or evaluation of another person at various dates, or it can be used in conjunction with any of the other sets. The information obtained reflects the subject's view of himself or the sorter's view of another person.

Typical applications of the Intrapersonal Test are:

Self-evaluation at various dates.	The "I" test. Repeated "I" tests.
Evaluation of alter at various dates.	The "He" test. Repeated "He" tests.
Comparison of two personalities.	Correlation of the "I" test with various "He" tests, of the "I" test with various other "I" tests, and of the "He" test with various other "He" tests.
Public opinion about one person.	Multiple "He" tests derived from various people wherein "he" is always the same person.

3. *The Interpersonal Test*

The Interpersonal Test consists of two sets of 100 cards each (see Table 3) on which are typed statements bearing upon the interaction between two persons. The subject is instructed to think specifically of his interaction with the designated other person when sorting the "I-him" and the "He-me" series. In other words, the "I-him" and the "He-me" sets are sorted by John Doe with regard to Tom Roe and by Tom Roe with regard to John Doe. The test can be used alone or in conjunction with the other sets. The analysis of an interpersonal relation is based on two persons' reciprocal sortings of the "I-him" and the "He-me" series. The four sortings obtained can then be correlated as follows.

TABLE 3
The Interpersonal Test: "I-Him" Version*

AB- 1
↓ See Intrapersonal Test and substitute there the word "him" for "myself."
63

AB- 64 I have a lot in common with him.
65 I avoid him when possible.
66 I learn from him.
67 I do things with him that I wouldn't dare do alone.
68 I antagonize him.
69 I joke with him.
70 I confide in him.
71 I am loyal to him.
72 I take advantage of him.
73 I feel equal to him.
74 I feel superior to him.
75 I am envious of him.
76 I am condescending towards him.
77 I express myself clearly with him.
78 I am apologetic to him.
79 I am sympathetic with him.
80 I keep him at a distance.
81 I am subtle with him.
82 I interrupt him.
83 I don't know how to behave with him.
84 I am humorous with him.
85 I coöperate with him.
86 I am stubborn with him.
87 I am original and imaginative with him.
88 I am submissive to him.
89 I resent the way he treats me.
90 I try to impress him.
91 I am responsive to him.
92 I try to get him to like me.
93 I am bothered by the way I behave (or feel) towards him.
94 I give him my full attention.
95 I try to get his sympathy.
96 I let him take advantage of me.
97 I try to be like him.
98 I am overly confidential with him.
99 When with him, I feel inadequate.
100 I make him angry.

---

*For the "He-Me" version, substitute the words "he" for "I" and "me" for "him."

Comparison of:

"I-him"^{A} with "I-him."^{B}	Reveals agreement about similarities in expectancies about self and about social actions; reveals the processes of identification.
"He-me"^{A} with "He-me."^{B}	Taps essentially the expectancies about the other person and the processes of projection.

"I-him"^{A} with "He-Me"^{A} or "I-him"^{B} with "He-me."^{B}	Sheds light upon a person's discrimination and differentiation of social action according to personalities. It is indicative of the awareness of ego-boundaries.
"I-him"^{A} with "He-me"^{B} or "I-him"^{B} with "He-me."^{A}	Checks the agreement between two people about social actions and their effects.

Typical applications of the Interpersonal Test are:

Evaluation of two-person teams Doctor-patient relationships. Mother-child relationships. Marital relationships. Friendship between two people. Rivalries between two people.	Correlation of the "I-him" and the "He-me" tests.
Comparison of several two-person teams The variety of an individual's social selves	Correlation of the "I-him" and the "He-me" tests.
Analysis of larger groups by analyzing two-person coalitions Detection of coalitions and the systematics of interpersonal shifts.	Correlation of the "I-him" and the "He-me" tests.

4. *The Intragroup Test*

While the Interpersonal Test attempts to examine the relationship of two people without reference to the wider social structure, the Intragroup Test examines the relationship of people within their social context.

The test material consists of two sets of 100 cards each (see Table 4) on which are typed statements referring to internal functioning of a group. The two versions consist of the "We" set and the "They" set and are sorted by the subject with a designated group in mind.

Typical applications of the Intragroup Test are:

Evaluation of group structure

Opinions of membership about the own group.	"We" test sorted by all members.
Opinions of membership about the out group.	"They" test sorted by all members.

TABLE 4
THE INTRAGROUP TEST: "WE" VERSION*

C- 1 ↓ 63 See Intrapersonal Test and substitute there the words "we" for "I" and "each other" for "myself."

C- 64 ↓ 85 See Interpersonal Test and substitute there the words "we" for "I" and "each other" for "him."

C- 86 Our leader is dictatorial.
87 In our group, nobody wants to be the boss.
88 We never question a decision of our boss.
89 When we feel threatened, we stand united.
90 Our group makes a smooth running team.
91 We are self-sufficient.
92 Our morale is high.
93 Our group is inactive or indifferent.
94 In our group, each of us has his own task well defined.
95 Within our group we can all substitute for each other.
96 Ours is a self-governing group.
97 We have different kinds of people in our midst.
98 Usually we all speak up at the same time.
99 In our group there exists a strong pressure for all to agree.
100 We are an enterprising group.

---

*For the "They" version, substitute the words "they" and "their" for "we" and "our."

Comparison of small groups.	Composite of "We" test compared to "They" test. Composite of "We" and "They" tests compared to other "We" and "They" tests.
Comparison of group opinion at various dates.	Repeated "We" tests and repeated "They" tests.
Determination of greatest or smallest weight of an individual's opinion in group decisions.	Comparison of individual sortings with ratings obtained by group majority vote ("We," "They," "We-them," and "They-us" tests).
Determination of coalitions between several people against other coalitions in the group. Analysis of leading ideologies.	Correlation of individual sortings and comparison of clusterings.
Comparison of an individual as individual to an individual as a member of a group	Comparison of the "I" with the "We" and the "He" with the "They" tests.

5. *The Intergroup Test*

The Intergroup Test consists of two sets of one hundred cards each (see Table 5) labelled the "We-them" and the "They-us" tests. The Intergroup

Test corresponds to the Interpersonal Test and is pertinent for the analysis of groups and people who frequently deal with other groups. While the Intragroup Test bears upon group structure, the Intergroup Test bears upon the social action of groups. When interaction between two groups is minimal, information can still be derived about the subjects' or whole group's expectancies (projections, prejudices, attributions).

TABLE 5

THE INTERGROUP TEST: "WE-THEM" VERSION*

CD- 1 ↓ 63 See Intrapersonal Test and substitute there the words "we" for "I" and "them" for "him."

CD- 64 ↓ 85 See Interpersonal Test and substitute there the words "we" for "I" and "them" for "him."

CD- 86 When we are with them, one of us is spokesman for our group.
87 We unite with them in a common goal.
88 We never discuss our internal problems with them.
89 We give them plenty of rope.
90 Whatever they do to us, we do to them.
91 Our way of doing things is different from theirs.
92 We try to break them up.
93 We are independent of them.
94 We dislike their way of doing things.
95 We are proud to know them.
96 We are unfair to them.
97 We try to win them over.
98 We upset their unity.
99 We comply with their requests.
100 We agree with their criticism.

*For the "They-Us" version, substitute the words "they" and "their" for "we" and "our," and "us" and "our" for "them" and "their."

Typical applications of the Intergroup Test:

Evaluation of group interactions.	Composite "We-them" compared to composite "They-us."
Evaluation of prejudice.	Comparison of "We" and "We-them" with "They" and "They-us" tests.
Evaluation of changes in group relations.	Comparison of "We-them" and "They-us" tests at various dates.

6. *The Communication Test*

The Communication Test does not form an integral part of the previously described test battery. It stands by itself and is thought of as a tool which enables a suitable sorter to summarize his impressions of an individual in terms relevant for a theory of human communication.

The test consists of 100 selected statements (see Table 6) which describe an individual's habitual communicative behavior. The statements are intended to be used by trained observers—that is, psychiatrists, psychologists, and anthropologists. The language used is somewhat technical and presupposes some knowledge of human communication theory (5). The cards are Q-sorted as in the other tests. The Communication Test is not a reciprocally sorted test, inasmuch as the patient or the lay person cannot be asked to sort the cards. The personal biases of the individual sorter are accounted for by using several trained observers and averaging their impressions. Adequate reliability is obtainable.

Because of differences in the universes being tapped, the Communication Test cannot be directly related to the other tests. However, relationships are discernible by utilizing the principle of stratifying subjects in terms of one test, e.g., Communication, and ascertaining the concomitant effect of this stratification upon the other tests.

7. *Applications of the Method*

As described earlier, the interrelation of levels is a feature of the four tests which increases greatly the kinds of information available. The following combinations of tests are of particular interest.

1. Comparison of the Intrapersonal Test with the Intragroup Test provides an indication of how an individual's picture of himself shifts when he views himself alone as compared to himself as a member of a group.
2. Comparison of the Intrapersonal Test with the Interpersonal Test yields information about the differences between the individual's perceived "real" self and his various social selves.
3. Comparison of the Intragroup Test with the Intergroup Test taps the ingroup-outgroup relationship and is relevant for understanding the nature of group membership.
4. Comparison of the Interpersonal Test with the Intragroup Test yields information about the individual's intimate, two-person functioning and his functioning in a group.

A number of specifically psychiatric problems may be investigated by the present method. The following are especially pertinent.

1. Evaluation of parent-child problems within the framework of the family.	Comparison of "I-him" ("I-her") and "He-me" ("She-me") tests sorted by both parents about each other and about the child. Where possible, the child sorts also.

TABLE 6
THE COMMUNICATION TEST

The Levels of Communication:

Habitual Preference of Levels

F- 1 Seeks to be alone with himself.
2 Seeks two-person situations or intimate gatherings of friends.
3 Seeks group situations—social gatherings, clubs, travel tours.

Distortions of Levels

F- 4 When alone, talks to himself.
5 In a two-person situation, behaves as if he were in a group: addresses other people as an audience.
6 When faced with a personal problem, behaves as if it had national importance.
7 When faced with a group situation, behaves as if he were alone; e.g., he keeps on talking without regard for others.
8 Makes group situations interpersonal by singling out one individual to talk to.

The Functions of Communication:

Preference of Functions

F- 9 When alone, likes to observe.
10 When alone, likes to mediate.
11 When alone, likes to be active and do things.
12 In a two-person situation, flexibly alternates between listening and talking.
13 In a two-person situation, acts as a listener or spectator.
14 In a two-person situation, takes the initiative; emphasizes talking or action.
15 Specializes in being leg-man or message boy for others.
16 In a group situation, is flexible and changes his behavior according to the needs of the situation.

Reception

F- 17 Is interested in self.
18 Is interested in things outside of himself.
19 Is interested in in-group functions: own family, team, firm.
20 Is interested in out-group functions: other families, teams, firms.
21 Is not satisfied with hearing or seeing, but has to touch, taste, and smell.
22 Is satisfied with inspection of things and people at a distance.

Evaluation

F- 23 Disregards past history of people and things.
24 Evaluates social behavior by what has been or can be achieved.
25 Evaluates social behavior according to popularity, public opinion, social pressure.
26 In evaluating social behavior, uses moral judgments.
27 Evaluates social behavior according to elegance and beauty.
28 Neglects long-term considerations in the evaluation of social behavior.
29 Evaluates social behavior according to impractical, idealistic, and unworldly principles.
30 Evaluates social behavior in terms of social progress.
31 In evaluating social behavior, cannot accept information from other people.
32 Uses first hand knowledge (participation) in evaluating group situations.
33 Oversimplifies information for the sake of action.
34 Acts upon relatively incomplete information.
35 Does not critically check the reliability of information.
36 Acts upon concrete information.

TABLE 6 (*continued*)

37 Tends to isolate himself socially (the "under-adapter").
38 Needs a lot of people around him whom he wishes to please (the "over-adapter").
39 Functions satisfactorily in well-defined situations only (the "over-controller").
40 Is constantly at odds with himself and his society.
F- 41 In a two-person situation, uses disagreement productively to arrive at a more correct viewpoint.
42 In a two-person situation, tries to convert others to his opinion.
43 In a two-person situation, when disagreement occurs, accepts the other person's opinion.
44 In a conflicting two-person situation, tends to break up the relationship.
45 In presence of value conflicts, favors change.
46 Is unaware of value conflicts.

Transmission

F- 47 Floods people with messages without regard for their receptiveness.
48 Pays attention to whether he has made himself clear to the other person.
49 His ambiguous means of expression elicits ambiguous responses.
50 Explains everything in detail.
51 Expresses himself by allusions and connotations.
52 In expressing himself, appeals to the emotions of others.
53 Makes ample use of facial expression and gesture in expressing himself.
54 Thinks and acts at a fast rate.
55 Verbal expression is predicate oriented (adjectives, adverbs).
56 His thinking and his verbal expressions are subject oriented (nouns).
57 In talking or writing, emphasizes action (verbs).
58 Is excessively verbal.
59 Expresses himself primarily through non-verbal means.
60 His means of expression appeal primarily to the eye and ear.
61 His means of expression appeal primarily to touch, taste, smell.
62 Utilizes autonomic nervous manifestations for purposes of expression.
63 Uses material goods to communicate with others: gives presents, exhibits luxury items, fails to return books.

Principles of Functioning

F- 64 Is method oriented.

Distortions of Functions

F- 65 Avoids making decisions.
66 Makes impulsive decisions.
67 Proceeds with action without forethought.
68 Tends to delay or avoid action.

Rôles:

Characteristics of Rôles

F- 69 Assumes rôles of equality.
70 Assumes rôles of inferiority.
71 Assumes rôles characterized by distance.
72 Is intolerant of people who are different from himself.
73 Behaves as if he were a fish out of the other pond (perennial stranger).

Rôles Adopted Inappropriately

F- 74 Treats everybody as if they were part of the family.

Character of Social Techniques

F- 75 Skilled in social techniques of intake.
76 Skilled in arguing and fighting.
77 Skilled in collecting, retaining and holding onto things and people.

TABLE 6 (*continued*)

78 Skilled in getting rid of things and people.
79 Tends to be ostentatious and exhibitionistic.
80 Skilled in the techniques of play, pretension, and humor.
F- 81 Skilled in the techniques of exploration and searching.
82 Tends to be passive and receptive.
83 Techniques geared to creating and conceiving the new.
84 Techniques geared to protecting, encouraging, and letting grow.

Rules

F- 85 Is a stickler for rules.
86 Is goal oriented.
87 Is a rebel—tends to violate rules.
88 Tends to be a reformer of rules.

Metacommunication:

Perception of Messages about Communication (Meta-Expectancy)

F- 89 Requires explicit instructions.
90 Tends to be unaware of own social clumsiness (oafishness).
91 Is capable of understanding allusions, hints, and other subtleties.

Transmission of Messages about Communication (Meta-Transmission)

F- 92 Possesses versatility in switching from one rôle to the other.
93 Tends to be a chameleon in assuming a variety of rôles.
94 Through assumption of different rôles, forces other people to adapt to him.

Feedback and Correction:

Individual's Method of Correction

F- 95 Capable of solving difficult problems for himself.
96 Relies upon others to correct his behavior.
97 Intolerant of errors in self and others.
98 Learns through trial and error.

Effects of Communication

F- 99 Is capable of communicating ideas clearly and efficiently.
100 Enjoys communication.

---

2. Evaluation of improvement of individual patient through therapy.	Repeated sortings of "I" and "We" tests.
3. Evaluation of effect of individual psychotherapy.	Repeated "I-him" and "He-me" ratings done by both doctor and patient.
4. Evaluation of effect of group therapy.	Repeated sortings of "I" test and "We" test and correlation of these tests.
5. Evaluation of treatment when both partners are in treatment.	Repeated "I-him" and "He-me" ratings of the partners correlated with ratings of the respective doctor-patient teams.
6. Evaluation of group phenomena on wards and between staff members.	Multiple "I-him" and "He-me" sortings correlated with individual sortings of "We" test. Correlation of individual "We" sortings with each other.

A presentation in detail of the analytical procedures that may be used on the data provided by the four tests is beyond the scope of the present paper and is better provided by an actual presentation of worked-through results (1). However, an outline of the various ways in which data analysis may proceed is relevant. Basically four approaches are suggested.

1. *Correlation analysis.* Any two tests can be related by the product-moment method of correlation. Thus the "He" series can be correlated with the "They" series, or the various "We"s may be intercorrelated. Direct inspection of correlation groupings and of comparative correlation strengths can be used to suggest and test hypotheses. Correlation between levels is, of course, limited to those statements which are paired.

2. *Factor analysis.* Reduction of the data by means of factor analysis may be done in order to bring out the relationships too intricate and complex for the human brain to perceive. The method, however, is tedious and cumbersome and may, on occasion, be misleading.

3. *Item analysis.* Straightforward analysis of the significantly discrepant items is extremely informative. Both in its own right and in conjunction with factor analysis, the method is fruitful and relatively simple. Of course, the problem of interpreting the significance of the items still remains.

4. *Construct indices.* On occasion, certain combinations of data suggest themselves as relevant transforms or representations of theoretical constructs. The resultant indices thus may be used to test and ramify the consequences of the theory underlying the investigation.

9. *Variations in Test Procedure*

An average subject requires from 30 to 45 minutes for sorting 100 cards. Depending upon the purpose and the scope of the study, a subject may be required to sort a formidable number of sets, a procedure which may make an excessive demand upon the subject's good will. Therefore the experimenter may find it expedient to utilize a shortened version of the tests. For this purpose, the first 50 cards in each set may be used. They have been selected as having the greatest relevance and, of course, halve the labor involved. Category frequencies now become 1, 3, 6, 9, 12, 9, 6, 3, and 1.

D. SUMMARY

1. The present paper describes an approach to the assessment of social relations. The method consists of card sets aimed at representing various levels of communication. On these cards are typed statements representative of each level which the individual arranges according to the Q-sort procedure.

2. At each level of communication there exist two parallel tests:

Intrapersonal level	"I" set	100 cards
	"He" set	100 cards
Interpersonal level	"I-him" set	100 cards
	"He-me" set	100 cards
Intragroup level	"We" set	100 cards
	"They" set	100 cards
Intergroup level	"We-them" set	100 cards
	"They-us" set	100 cards

3. The innovation introduced with this test consists neither in the type of statements used nor in the method of sorting, but in the fact that reciprocal sortings between people and groups are compared with each other. The degree of agreement and the areas of disagreement furnish an index of the type and efficiency of the communication which took place in previous interactions.

4. Except for the pronouns, the wording of the statements throughout all the card sets has been kept similar in 63 out of 100 statements, so that product-moment correlations between levels of communication can be used in the analysis of the data.

5. An additional sorting method, consisting of 100 cards and describing the behavior of one person in terms of a theory of human communication has been compiled. It is sorted by trained observers and may be used to test hypotheses drawn from the theory.

6. The test method is applicable for the study of the single individual, two-person systems, group structure, or inter-group relations in such contexts as the study of marital relations, doctor-patient relations, family structure, labor-management relations, interaction between teams, and many others. The method as a whole is flexible and versatile in its applications. Any of the tests may be used singly or in combination with others, in its full length or in abbreviated form.

REFERENCES

1. BLOCK, J. The assessment of communication: II. Rôle variations as a function of interactional context. *J. Personal.* (In press.)
2. DEUTSCH, K. M. Mechanism, teleology, and mind: The theory of communications and some problems in philosophy and social science. *Phil. & Phenomenol. Res.,* 1951, **12**, 185-223.
3. FESTINGER, L., BACK, K., SCHACHTER, S., KELLEY, H. H., & THIBAUT, J. Theory and Experiment in Social Communication. Ann Arbor: Research Center for Dynamics, Institute for Social Research, Univ. Michigan, 1952. Pp. 123.
4. MILLER, G. A. Language and Communication. New York: McGraw-Hill, 1951. Pp. 298.
5. RUESCH, J., & BATESON, G. Communication, the Social Matrix of Psychiatry. New York: Norton, 1951. Pp. 314.
6. RUESCH, J. The therapeutic process from the point of view of communication theory. Paper presented at the symposium on "The Therapeutic Process" at the 29th annual meeting of the American Orthopsychiatric Association, Atlantic City, Feb. 26, 952. (In press.)
7. SMITH, B. L., LASSWELL, H. D., & CASEY, R. D. Propaganda, Communication, and Public Opinion. Princeton: Princeton Univ. Press, 1946. Pp. 435.
8. STEPHENSON, W. A statistical approach to typology; the study of trait universes. *J. Clin. Psychol.,* 1950, **6**, 26-38.
9. WIENER, N. Cybernetics, or Control and Communication in the Animal and the Machine. New York: Wiley, 1948. Pp. 194.
10. ————. The Human Use of Human Beings—Cybernetics and Society. Cambridge: Houghton Mifflin, 1950. Pp. 241.

University of California Hospital
University of California Medical Center
San Francisco 22, California

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 81-105.

STATUTORY DEFINITIONS OF FEEBLEMINDED IN U. S. A.*

Juvenile Court, Honolulu, Territory of Hawaii

STANLEY D. PORTEUS AND GERALD R. CORBETT

Interest in those unfortunates who suffer from severe mental handicaps has been steadily growing for very many years, yet the distinction between the mentally deficient and the mentally disordered is, in many people's minds, not at all clear. Consequently the fact that the one group constitutes mainly a psychological problem, the other a psychiatric problem, is often not recognized. This confusion may be largely due to the fact that the insane behave as though they were "bereft of their senses," to use a once popular descriptive phrase, whereas the mentally defective also frequently show disturbances of conduct. In defining feeblemindedness, some American jurisdictions have attempted to draw a legal distinction between the two groups of mentally incompetent by specifying that the condition of feeblemindedness, or arrest in mental development, must have become apparent from birth or at an early age. On the other hand, insanity or psychosis implies a breakdown in what was once normal functioning. The feebleminded never were socially or mentally efficient; the psychotic once were, but are so no longer. Burt, the English authority, points out that as early as the reign of Edward I the idiot was described as a "born fool," the lunatic one who "at birth hath had understanding, but by disease, grief, or other accident, hath lost the use of his reason."

To illustrate this difference, an analogy may be found in a comparison of two motor vehicles. One may be so built that its engine is of such low horsepower that it cannot climb ordinary hills or carry any but the lightest load. It represents the feebleminded. On the other hand, another automobile may be equipped with an engine of sufficient horsepower to function adequately but may lose its efficiency either by reason of unusual strains or stresses causing partial breakdown, or even through bad driving or the owner's neglect. The same analogy may be used to point the difference between the rôles of the psychologist and the psychiatrist. The former is concerned largely with detecting the faults or weaknesses in construction which render the car inadequate for average transportation requirements; the latter is concerned with keeping the car on the road, through the avoidance

---

*Received in the Editorial Office on August 26, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

of unusual conditions of travel, and the use of proper lubrication and fuel. The psychiatrist may sometimes effect general or specific repairs, though his work is mainly preventive. In the present state of our knowledge, feeblemindedness is not susceptible to repair. The best the psychologist can do is to make the road as easy as possible for the car.[1]

That there is considerable confusion in the understanding of the two conditions is evinced by the fact that it is often very difficult to find in the indexes of legal texts and compilations of statutes the references proper to the feebleminded. The laws which relate to them may be listed under "mentally incompetent persons," "mental defectives," "mental deficients," "mentally handicapped," or more frequently under the heading "insane," and only in a few instances under "feebleminded." In some jurisdictions the statutory provisions for definition, commitment, etc., are to be found under the name of the state training school, hospital, or institution where feebleminded are placed. A first step towards lessening confusion would be general adoption of the term "feebleminded." It is usually hyphenated as a compound word, but it is now of such general application that it would be well to use it as a single word without the hyphen.

A second source of confusion is that three different degrees of mental deficiency or feeblemindedness are distinguished by the separate terms of idiot, imbecile, moron or high-grade feebleminded. Some American states go so far as to incorporate definitions of these degrees of feeblemindedness into their laws. This practice seems quite unnecessary, although it follows English precedent. For purposes of institutional treatment the division of patients into idiots, imbeciles, and morons is useful, but in legal practice the use of the inclusive term "feebleminded" is sufficient.

In England, definition by degrees of mental subnormality was necessary because there the term feebleminded was reserved for defectives of high grade, or what are now frequently called "morons," a name devised by Dr. H. H. Goddard and which has now entered into common usage.[2] Because of this difference in terminology, it was thought necessary in England to define these mental subdivisions. According to the English Act of 1927, idiots are "unable to guard against common physical dangers"; imbeciles are "incapable of managing themselves or their affairs, or of being taught to do so"; the feebleminded are those "who require care, supervision and control for their own protection or for the protection of others," or, in the case of children, "appear to be permanently incapable of receiving proper benefit from the instruction in ordinary schools." The same phraseology has been adopted in some American jurisdictions as has also the definition of subclassifications of idiot, imbecile, etc. The use of the inclusive term "feebleminded" makes the definition of subdivisions unnecessary.

---

[1] The fact that this paper is intended for a mixed professional group of lawyers, judges, psychologists and psychiatrists has necessitated simple psychological as well as legal formulations.

[2] It has also achieved dictionary status. Webster defines moron as "a moderately feebleminded person."

Apparently the substitution of the educational for the social criterion in the English law relating to feebleminded was an attempt to meet the difficulty that children have not had the opportunity to prove their ability to manage themselves or their affairs. For reasons to be discussed later, the shift from one standard or criterion to another seems inadvisable.

In America, the Association for the Study of Feebleminded, a group of institutional superintendents, psychiatrists, and clinical psychologists, attempted in 1912 to clarify the matter by defining the three grades of mental defectives in psychometric or mental measurement terms. The upper limit for idiots was declared to be 3 years' mental age, for imbeciles 7 years, for morons 12 years. The above limits recently formed the basis for the law governing commitment of feebleminded in one State (Ohio), and in Maine, where no legal definition exists, and possibly elsewhere, these limits have been adopted as the basis of admission to the State institution for the feebleminded. But in modern psychological experience, 12 years' mental age has been found to be far too high a figure to serve as the lowest level of normality. Even Goddard who first proposed this standard has changed his position, suggesting 7 years' mental age as the possible limit of feebleminded development. This would be nearer present-day psychological opinion in England. Sir Cyril Burt (2) gives 8 years' mental age as being the upper limit for the feebleminded. The above limits all relate to "adults" but since in psychometric practice the highest actual age used in calculating *IQ*'s should be 14 years, the standard applies to all individuals over that age.

The 1943 edition of *Outlines for Psychiatric Examinations,* revised by Dr. Nolan Lewis (7) follows the same plan of definition by the psychometric level, expressed as an intelligence quotient (*IQ*). This is the mental age level by a standardized test, divided by the chronological age of the child. For an adult or adolescent, the divisor formerly used was 16 years, based on the belief that this was about the average test age of the community at large. The Army examinations of 1917-18 showed that this was too high, the average test age being a little below 14 years. The latter figure should therefore be used as the divisor, as only by such a procedure could the average *IQ* equal 100. In modern practice, the 16-year divisor used in calculating the *IQ* of anyone above that age has been abandoned in favor of a lower figure, the effect being to raise the adult individual's *IQ*. Formerly a person testing 14 years would have an *IQ* of 87.5. Now, by using 14 as the divisor, it would be 100. Using too high an average mental age would thus make differential diagnosis of high grade feebleminded untrustworthy.

The Lewis manual states that the idiot, *if a child,* has an *IQ* of less than 20; the imbecile of from 20 to 49. The ceiling for feeblemindedness is tentatively set at 69, but the suggestion is made that this limit should be disregarded "where medical, social, and other facts indicate that the patient is mentally defective." It is therefore implied that a person may have an *IQ* above 69, limit not specified, and still be feebleminded. No account is taken of the far more common situation, that individuals with *IQ*'s below 69 are *not* feebleminded. In computing *IQ*'s of adults, the manual states that the mental age is to be divided by 16 years. Not only the differences in practice in calculating an *IQ,* as referred to above, but the uncertainties of the psychometric concept, show how unsuitable it is as the basis of a legal definition. There is also the dubious implication that an *IQ* may have one diagnostic value if the individual concerned is a child, an entirely different significance if he is an adult.

This unsatisfactory assumption results from a wide discrepancy between estimates of the incidence of feeblemindedness founded on school surveys and those of the adult or general population. Using the criterion formerly accepted by many psychologists that children below 70 *IQ* are feebleminded, authorities on the subject have commonly stated that 2 per cent represents the percentage of feebleminded in schools, though actual surveys have shown as high as 6 per cent of pupils below that standard. On the other hand, even the least conservative estimates of the number of adult feebleminded rarely place the figure above 1 per cent. Our own estimate would be somewhat less than one-half of one per cent, or 4 per thousand of the general population.

It would follow, therefore, because of this difference in estimates between 2.0 and .04 per cent, that either the feebleminded die off or get better as they grow older; or according to another, much more tenable hypothesis, that diagnosis using 69 *IQ* as the "floor" of normal intelligence is extremely unreliable.[3] The first of the three explanations of this discrepancy is too morbid, the second too optimistic, whereas the untrustworthy nature of the psychometric definition has been borne out by every follow-up study of "feebleminded by test." Baller (1), for example, studied the after-careers of 206 children declared by psychologists to be feebleminded on the basis of test levels below 70 *IQ*. Less than 7 per cent were in feebleminded institutions; 19 per cent had been in corrective institutions; but 56 per cent had been subjected to no special control or supervision and were apparently subsisting through their own efforts.

---

[3] Another factor in the situation is that the obviously feebleminded, i.e., those of idiot and imbecile grades are excluded from school. Adding this group would merely increase the percentage of children considered feebleminded and widen the discrepancy between that figure and estimates of the number of adult mental defectives.

If the estimate of 4 per thousand feebleminded in the population were accepted, that would mean about 600,000 mental defectives in the U. S. A. In 1942, inmates of feebleminded institutions numbered 105,000, and according to Weihofen (11), there are now "more than 110,000 mentally defective persons in American institutions." This would mean that at least four out of every five feebleminded individuals are at large in the community, and might indicate, either that the number of these unfortunates is still overestimated, or that the problem of the high-grade feebleminded in the community is not as serious as some social reformers would have us believe. In any case, the estimate of 2 per cent school children feebleminded or about three million defectives in the U. S. A. is obviously extremely dubious, and would provide a weighty reason why a legal definition based on psychometric limits should be avoided. This does not mean that psychometric levels have no significance, but what needs emphasis is the fact that the social implications of a subnormal test score are far from clear. Obviously there are very many people who score low on the tests who seem able to get along satisfactorily in society—in other words, are socially sufficient.

The publication of the results of Army examinations of recruits during World War I should have dissipated completely the idea that 70 *IQ* was the dividing line between normals and feebleminded. Using the method of calculating *IQ*'s in vogue at that time, 70 *IQ* corresponded to 11.2 years mental age. Having converted Army test scores into their equivalent Binet mental ages, some authorities calculated that the acceptance of this standard would mean that about thirty millions of persons in the U. S. A. could be classified as mentally defective. One would have supposed that such a ridiculous conclusion would have resulted in the instant and complete abandonment of the proposed standard. Nevertheless, social workers, psychiatrists, and psychologists are in very many cases using that standard today. This may in some measure account also for the fact that the number of inmates of feebleminded institutions rose from less than 5,000 cases in 1890 to over 110,000 in 1952.

The widespread use of mental tests in general has undoubtedly affected the situation. For example, a survey conducted by Dr. Lewis of England in 1926 gave the incidence of serious mental defect as 8.57 per 1,000, just about double a previous estimate made in 1904. Figures for the U. S. founded upon Draft Board experience in World War I have been stated as 9.24 per thousand. The injudicious use of psychometric levels in the diagnosis of feeblemindedness has undoubtedly resulted in the segregation of a large number of individuals who are definitely socially inadequate, but not necessarily so because of mental deficiency. For example, a very recent survey of a large group of institutional cases in England by Tizard (10) revealed the fact that between 25 and 50 per cent of these cases would, according to this author's statement, be better classed as dull or subnormal than feebleminded, if properly conservative methods of diagnosis were employed. A similar survey of institutions in America would probably show a considerable percentage of inmates doubtfully diagnosed. Thus the assumption commonly made in many states of the Union that the responsibility for certifying individuals to courts for commitment can be safely left to physicians and psychologists and that therefore there is no need for a legal definition of feebleminded, is not sound. An illustration of this assumption is afforded in an official letter describing the situation in a southern state. The writer says: "The law covering feeblemindedness or mental deficiency . . . makes no attempt to define the condition but appears to assume that it is already known, or that doctors and educators will be well informed on the subject."

A major factor in the upsurge of interest in the care and control of feebleminded was the publication of Dr. Goddard's book *The Kallikak Family* in 1912 (4). It sets forth the lines of descent from the casual union of a Revolutionary officer and an alleged feebleminded girl. Just prior to Goddard's study of heredity, a wave of interest in the newly devised instrument of mental tests had resulted in general surveys of corrective and other institutions, and the application of the too-severe standards of normality already commented upon, seemed to show very large percentages of inmates feebleminded. Field workers, influenced by such reports, did not hesitate to mark down on their charts as defective anyone with a bad criminal, delinquent, sexual, or social record. Alcoholism or having an illegitimate child was considered sufficient proof of feeblemindedness. Consequently the Kallikak family tree was loaded with mental defectives, in most remarkable contrast with the absence of mental taint in the respectable branch of the Kallikak family which stemmed from the later marriage of the officer with a normal woman of good family. Of 480 descendants of the alleged defective girl, 143 were put down as feebleminded; of 496 descendants of the respectable marriage, all were stated to be normal, though three were "somewhat degenerate," two of whom were "alcoholic," and one "sexually loose."

As a result of these very questionable findings, a strong campaign was set on foot to induce state legislatures to make better provision for the feebleminded. The demonstration was so convincing that law-makers were easily persuaded that feeblemindedness was the most acute social problem and that checking the reproduction of the mentally unfit would have the effect of lessening crime, pauperism, prostitution, alcoholism, and other social evils. Some states passed sterilization laws, which usually became inoperative, while many others passed laws for the segregation of the feebleminded. Apart from the statutes providing for sterilization of defectives, some reminders of the campaign are still to be found. Alabama, for example, tacked onto its law governing the definition and commitment of feebleminded this pronouncement: "It is specifically recognized that the greatest danger that the feebleminded constitute to the community lies in the frequency of the passing on of mental defect from one generation to another" (Code of Alabama, 1940, Title 45, Sec. 236).

South Dakota also states that the purpose of its law on the subject is to make the feebleminded wards of the State "and segregate them to the end they should not reproduce their kind" (South Dakota Code of 1939, Sec. 34.0401).

Perhaps this whole matter can be summed up best by saying that though the results of the Kallikak Family study were decidedly questionable, its publication did stimulate a very worthwhile movement in the direction of better care and provision for the feebleminded.

One consideration that profoundly affects this segregation policy is the permanence, or the incurability, of feeblemindedness. Despite the wide discrepancy between the estimated incidence of the condition among adults as compared with those for children, and the occasional publication of optimistic reports about progress under special systems of education, or even special medication such as glutamic acid feeding, the feebleminded are not, in the present state of medical or psychological knowledge, curable. Nor do surgical methods avail in mental deficiency as they have been recently applied for the "cure" or removal of certain psychotic conditions. This concept of incurability illustrates one essential difference between psychotic cases and the feebleminded. As for special training, various methods have been tried since the time of Itard, a French physician-psychologist who in

1804 made a most patient and intensive effort to educate the so-called "wild boy of Avignon," only to discover that his patient was not an untutored waif of the forest, but a mental defective, and ineducable. A century and a half after Itard's effort, we are no nearer an educational solution of the problem, in spite of widely heralded systems such as that of Seguin, and later of Montessori at the turn of the century. Latterly there have been short-lived and much more dubious claims.[4] It seems certain that the feebleminded cannot be made socially wholly independent, any more than the dull can become bright, or the man of ordinary talent can rise to genius. Whenever such developments appear to have occurred, the only safe conclusion is that the former mental diagnosis was a mistaken one. Very rarely, one meets with socially defective individuals with specialized or one-sided mental abilities such as capacity for instantaneous arithmetical computations, phenomenal memory, etc., these persons being referred to in the literature as "idiot savants."

We have already noted that the English definition of an imbecile is one who is incapable of managing himself or his affairs, *"or of being taught to do so,"* thus emphasizing the incurability of the condition. In America, the tendency has been to copy the English law, so that California, Illinois, Louisiana, Montana, Nebraska, Virginia, all declare in their statutes that the feebleminded are incapable of managing themselves socially, "or of being taught to do so." However, Montana's law seems self-contradictory, since it uses the above phraseology in describing persons eligible for admission to its State Training Schools, and then goes on to include "minors whose intelligence will not develop without such care"—which certainly suggests that institutional training may cure the condition (Laws of Montana, Chapter 183-5, as quoted in an official letter from the Montana State Training School). That institutional training can increase intelligence would hardly be admitted if the usual concept of intelligence as "inborn or innate ability" is accepted.

While this paper is directly concerned with statutes relating directly to the commitment of mental defectives in a proceeding designed exclusively for that purpose, our comments have a direct bearing upon statutes relating to guardianship for incompetent persons and statutes relating to the commitment or transfer to institutions for mental defectives of persons accused or convicted of crime.

In making this survey of the laws on the subject, we have been struck by the remarkable fact that though all American jurisdictions make some provision for the segregation of the feebleminded, 20 or 40 per cent of the total have no statutory definition of the condition. This is remarkable because commitment to an institution may be for life, and such a deprivation of liberty should certainly be hedged about with every possible safeguard against unjust or mistaken commitment.

---

[4] See a recent *Reader's Digest* article "Feeblemindedness Can Be Cured."

In addition, and even though long term institutionalization may not result, any official pronouncement of feeblemindedness is bound to affect drastically the entire course of the life of the individual concerned; to limit—or eliminate—his right to contract in relation to his services or his property, his right to marry, his right to dispose of his property by will, and many other rights enjoyed by the "normal" citizen.

However definitions of feeblemindedness differ, they seem to agree that it is a condition of social inadequacy dependent upon retarded mental development. To judge of the degree of mental deficiency makes it necessary to determine the stage of arrest of development. This involves mental measurements and the correlation of these with social capacity so that in the case of children a prediction both of the ultimate mental development and its relation to social competency is possible. Such mental and social diagnoses can only be made by experienced clinical psychologists and psychiatrists who are fully aware of the limitations of those mental measurements, particularly in relation to the individual's self-management and self-support. To leave the determination of the condition to two or more physicians, as is done in many jurisdictions, and without the guidance of a statutory definition with which the results of their examinations must conform, is to deprive the alleged feebleminded person of an essential protection. Or are we to assume that the feebleminded are not entitled to the guarantees of our Bill of Rights? Even if the medical examiners are honest and of excellent reputation (a proviso which lawmakers in some states thought necessary to insert in their statutes), their lack of diagnostic experience with feebleminded of moron grade may mean that only obvious low-grade cases are committed, whereas the community's greatest concern may be with the segregation of the high-grade defective.

As far as we can ascertain from our sources of information, jurisdictions without statutory definitions are Alaska, Arkansas, Connecticut, Delaware, Florida, Indiana, Kansas, Massachusetts, Maine, Maryland, Michigan, Missouri, North Carolina, New Jersey, New Hampshire, Nevada, Oregon, Rhode Island, Vermont, and Wisconsin, their geographical range covering all the principal sections of the U. S. A.

Under the plan adopted for the survey herein reported, the investigators, after using whatever resources were available in the library of the Supreme Court of Hawaii, and realizing that in some instances our references may not be up to date, addressed letters to medical superintendents of institutions in the above-named States which were listed as having no statutory definitions, and judging by some of the replies received, the lack of such a definition is keenly felt. In some cases the superintendents have been constrained to set up or adopt their own standards for admission to their institutions. Sample comments are as follows:

> "We do not find anything that even remotely resembles a legal definition of feeblemindedness that forms the basis of commitment. This is rather startling, but it is the truth."
>
> "The law (of this State) makes no attempt to define the condition, but appears to assume that it is already known, or that doctors and educators will be well informed on this subject."
>
> "I am sorry to say that the state laws use the word 'feebleminded' but do not define it. . . . Since we have no legal definition in the state we have been obliged to accept as a basis for diagnosis the definition of feeblemindedness as adopted by the American Association on Mental Deficiency." (This definition uses as a criterion social incapacity resulting from arrested or imperfect mental development. This development, in morons, "does not exceed that of a normal child of about twelve years.")[5]
>
> "In (our State) mental deficiency is coupled with insanity, so we have rather a mixed-up mess. Indeed, we don't have any really clear-cut legal definition."
>
> "The only definition of 'feeblemindedness' seems to be in Section 51.001 where it states that 'mental deficiency' is synonymous with 'feeblemindedness.'"
>
> "You are doing a great service in attempting to bring some type of order to the legal definitions which we now have."
>
> "Such a survey would be of great value toward modernizing the antiquated laws pertaining to this subject that now exists in our state statutes."

From the above typical comments, it would seem that medical men and psychologists most closely in touch with the problem of the feebleminded would welcome a clarification of the subject that might result in the adoption of a satisfactory definition.

Some states which do have statutory definitions are apparently trying to define the condition too quantitatively by putting their faith in mental test standards that have now been discarded by competent psychologists. It would be very convenient if it were possible to apply a test such as the Binet or the Wechsler-Bellevue, work out an intelligence quotient, and draw an exact line between normal and feebleminded, either in terms of a quotient or a given mental age.

---

[5] For the *IQ* limits proposed by this Association, see Lewis, *Outline of Psychiatric Examinations,* previously quoted.

Such a dividing line was once proposed by Dr. Terman (9), who devised the Stanford-Binet Scale and later published the Terman-Merrill Revision. Writing in 1917, he declared: "All who test below 70 *IQ* by the Standard Revision of the Binet-Simon Scale should be considered feebleminded, and it is an open question whether it would not be justifiable to consider 75 *IQ* as the lower limit of normal intelligence." But in modern clinical practice few psychologists believe that such a definite dividing line can be drawn. Many authorities on the subject of feeblemindedness, such as Doll (3), have definitely declared themselves against this idea. Goddard (5), who in 1912, as previously mentioned, had proposed a limit of 12 years' mental age in relation to high-grade feebleminded (for whom he invented the term moron), in later years also questioned the validity of this standard. Writing in 1928, he said: "Is it possible that during all these years we have placed the limit of feeblemindedness too high? Is the real limit 7 years instead of 12?" The right answer to the question is—neither limit is correct. No single scale of tests that the psychologist uses can be wholly relied upon for final diagnosis. The task before the mental examiner is to re-value his tests, especially in the light of new knowledge that has accumulated in the past 10 years through brain surgery, and decide for himself which tests have the greatest predictive value in terms of social adequacy. By combining the verdicts of several of these tests, he would then not hesitate to call an adult individual testing below 9 years feebleminded, if his social or educational status was similarly low; but even with these reservations it is impossible to fix an exact dividing line that will segregate "the sheep from the goats." But there is little doubt that if the above suggestions were followed by examiners the borderline zone between normal and feebleminded would be narrowed and much more thinly populated. What we should like to emphasize is the inadvisability of attempting a statutory definition in exact psychometric terms. It is somewhat analogous to declaring that a definite blood pressure distinguishes a normal from an abnormal heart condition.

A further quotation from Dr. Goddard will serve to illustrate the present confusion of psychological ideas on the subject. "The morons," he says, "comprise that large group of people whom we recognize as of dull intelligence who never live sumptuously, or manage themselves or their affairs with a high degree of prudence, but nevertheless do manage themselves; and, if carefully and wisely trained, are not a burden on the community." This would mean that morons are not, in the social sense, feebleminded, unless they have missed being "carefully and wisely trained," presumably in institutions for the feebleminded. There is, of course, no guarantee that this training can or will be given.

Hence the courts would be likely to commit only those who test below moron status, i.e., those with mental ages of 7 years or below, which would result in lack of supervision and control for the many morons who would thus become "a burden on the community." Again, we would urge that the statutes should have no concern with subdivisions of mentality such as idiot, imbecile, and moron or high-grade feebleminded, or with the problem of determining diagnostic psychometric limits or standards, but should leave this matter to the experts, with whom it belongs.

The shift in psychometric standards is well illustrated by changes in the statutes dealing with feebleminded that have taken place in Ohio, one of the few states that have set up this kind of criterion At one time the law read as follows: "The term 'feebleminded' means the condition of any person found within the State of Ohio, having a mental age of 12 years or less, and having an intelligence quotient of 69 or less. . . ." The statute then went on to define an idiot as having an *IQ* of less than 25, an imbecile as having an *IQ* of not less than 25 or more than 50, a moron as having not less than a 50 *IQ* or more than 69 (Throckmorton's Ohio Code Annotated, 1940, Sec. 1890-97).

An amended law, effective in 1941, dropped all reference to the mental age limits of mental deficiency and set up the following definition which combines the social and psychometric standards: "The term 'feebleminded,' or mentally defective, refers to any person who is not mentally ill, but whose intellectual development is so retarded that he has not acquired sufficient control, judgment, and discretion to manage himself and his affairs, and for whose own welfare or that of others, supervision, guidance, care, or control are necessary or desirable" (Throckmorton's Ohio Code Annotated, 1940-45, Sec 1890-97).

Still later, an Ohio statute defines a mentally deficient youth as a person under 21 years of age, "incapable of profiting substantially by any educational program" provided by a public school, or if not of school age, having an *IQ* below 50. The Commissioner of Mental Hygiene is given the responsibility of prescribing the manner of determining the degree of mental deficiency (Throckmorton's Ohio Code Annotated, 1948-52, Sec. 1890-7a).

The shift in emphasis from a more definitely restricted and more severe criterion to a much lower psychometric standard should be noted in comparing the above mentioned Ohio statutes. It is also interesting to observe incorporation into the definition of the social criterion of incapacity for self-management, and later the reference to the educational criterion of inability to profit by instruction in the public schools because of mental deficiency.

Oklahoma also by implication sets up a psychometric definition, but one which is more liberal than Ohio's. The law governing admission to the Enid State School (for mental defectives) classifies those to be admitted as follows:

> (1). White children (under 21 years of age), "with idiocy, imbecility, or low moronic level."
>
> (2). White adults (21 years or older), of either sex, "with a mental defectiveness equivalent to approximately the intellectual level of a nine-year child or less." (Oklahoma Statutes 1951, Title 43-A, Sec. 30).

No doubt in deference to the fact—whether legally recognized or otherwise—that diagnosis of feeblemindedness is now carried on everywhere with the aid of mental tests, other jurisdictions imply such usage in their statutes. Virginia, for example, defines (somewhat unnecessarily) an idiot as "a person who from birth or by reason of failure of early development has not attained a maximum mental age of three years according to the Binet or other approved mental test, and who is incapable of making known his actual wants, to feed himself, and is practically deficient in the mentality of a normal child of three years of age."

Included in the same law is a good, fairly comprehensive definition of a feebleminded person, with no mental age limits stated. Another section provides that any individual admitted to the State Institution "is to be subjected to the Binet-Simon measuring scale for intelligence or some other approved test of mentality" (Code of Virginia, 1950, Sections 37-176, 37-177, 37-193, 37-210).

Still other States indirectly recognize psychological criteria by providing that a psychologist be included among the examiners who certify feeblemindedness to the court. At one time in Arizona the examining board consisted of one physician with at least three years' practice, one psychiatrist, and one psychologist. This law was repealed and a later law provided that "the examiners should be two reputable physicians who shall use standard mental and psychological tests" to determine that the child will benefit by the Arizona Children's Colony Services. Section 15 of the same law provides that the superintendent of the Colony should be a reputable trained psychologist with five years' administrative institutional experience (Session Laws of Arizona, 1949, Chapter 30, H.B. 62).

In California, examination must be at the hands of a physician who has made a special study of mental deficiency, and a clinical psychologist, or by two such physicians or two such psychologists (Deering's Welfare and Institutions Code, Section 5257).

The Delaware statute requires the committing judge to hear witnesses "one of whom shall be a psychologist, or an expert on feeblemindedness" (Revised Code of Delaware, 1935, Section 2611).

Georgia, though it does not include a psychologist on its board of "two physicians possessed of knowledge and training in mental diseases," does require these physicians to make "complete examinations both physical and mental" of the alleged feebleminded case. Obviously a mental examination would hardly be complete if tests were not used to determine the fact of "mental defectiveness" on which basis individuals are admitted to the State Training School (Code of Georgia, 1933, Sec. 35-304).

Iowa makes a similar provision of one physician and one psychologist, but adds that the commission must make a personal examination, i.e., testing cannot be deputed to others (Code of Iowa, 1950, Sec. 222-14). Minnesota makes no provision for a psychologist on its commission of two licensed doctors of medicine, but adds to this examining board "one person skilled in the ascertainment of mental deficiency" (Minnesota S. A. 1944, 525-749).[6]

Nebraska follows the same rule of providing that a psychologist may be employed to assist the Commission for the Control of Feebleminded (Revised Statutes of Nebraska 1943 (re-issue of 1950) Chap. 83, Art. 228). Oregon also requires examination by two or more competent physicians or by two or more competent physicians and psychologists (Oregon Compiled Laws, 1940, Sec. 127-303). Virginia goes so far as to state the qualifications of the examining psychologist as "any person, male or female, skilled in making Binet-Simon or other approved mental tests, and certified as a certified clinical psychologist by the Commission of Mental Hygiene and Hospitals" (Code of Virginia 1950, Sec. 37-193). West Virginia, like Idaho and Georgia, specifies that the examination must be "a complete and thorough examination, both mental and physical" (West Virginia Code of 1943, Section 2686). Wyoming requires *separate* examinations by a physician and a psychologist, the latter being either the State Director of Special or the Superintendent of the State Training School, or other person designated by either of these two as competent (Wyoming Compiled Statutes, 1945, Sec. 19-420).

---

[6] See also a later reference in this article to a Minnesota Supreme Court decision in the case of a woman previously declared feebleminded and committed to an institution.

In Illinois, an examining commission is set up, consisting of two qualified physicians or one qualified physician and one qualified psychologist. The qualifications set forth for the latter are a master's degree in psychology, plus two years full time in the practice of clinical psychology, or one year of full time experience in the practice of clinical psychology with mentally defective persons. In our opinion possession of a doctor's degree in psychology would provide an additional safeguard (Illinois Revised Statutes 1949, Chapter 91½, Sec. 55-5).

In New York state, besides voluntary and court commitment, the law provides that an alleged mental defective may be admitted to a state institution "on certificate of one physician or one psychologist" on the basis of a "final" examination made within 60 days next before such admission (Mental Hygiene Law, 1939, Art. 6, Sec. 121).

Under this law there are apparently two possibly inadequate safeguards against wrongful segregation of an individual. He cannot be retained at the institution more than 30 days without further inquiry, if he or someone on his behalf makes a written request for release (Sec. 123). The other check is that the superintendent of the institution may refuse to admit or may release the individual if in his judgment the latter is not feebleminded. He also has the right to take this action if the person has been committed by court action. Thus the institutional superintendent is the final arbiter (Sec. 124-7).

But such an official, though an excellent administrator, may not be expert in mental diagnosis. In any case he would probably delegate examinations to a member of his staff who may be a psychologist of limited experience, or of inadequate understanding of the pitfalls in mental testing. There is a tendency among the staff of some institutions to regard admission as proof of mental deficiency. Moreover, the longer and more intimate the connection with institutional work, the more likely it is that the official is out of touch with ordinary community conditions. Unfortunately, a similar limitation of view often affects the psychologist if his contact with his cases is confined to office examinations. He is likely to become imbued with a blind faith in mental tests as predictive measures of social competency. Social myopia is common in many professions.

It is odd to note that New York state requires, in the case of court commitments, certificates from two qualified examiners or one qualified examiner and one qualified psychologist, whereas, when there are no court procedures, a certificate from one examiner suffices. With high-grade or borderline cases, diagnostic errors can and do occur, and it is sometimes much easier to get a person into an institution than to get him out. Hence the more checks or safeguards, the better.

In Hawaii, the definition is based on incapacity for self-management and self-support by reason of deficient mental development "as demonstrated by standard psychological tests." The statute also provides that the application for commitment to the feebleminded home must be accompanied by a certificate by a specially qualified physician or by a psychologist employed at the Territorial Psychological Clinic, or a psychologist certified by this clinic (R. L. Hawaii, 1945, Sections 4064-4066). Incidentally, there are many "standard" tests which have little diagnostic value; if the use of tests is prescribed by statute, these tests should have proved validity in the diagnosis of mental defect.

Iowa also provides for the appointment by the court of a commission to consist of two qualified physicians, or one qualified physician and one qualified psychologist, to examine and determine the mental condition of an alleged mental defective (Code of Iowa, 1950, Chapter 222-14).

In summary, the jurisdictions that, either by means of a psychometric definition or indirectly by specifying mental tests in an examination, recognize the need for psychological examinations, number nineteen. They are Arizona, California, Delaware, Hawaii, Georgia, Idaho, Illinois, Iowa, Louisiana, Minnesota, Ohio, New York, Oklahoma, Oregon, Nebraska, New Hampshire, Virginia, West Virginia, and Wyoming. But the heads of institutions in both Kansas and Maine, states without statutory provisions, set up mental standards for admission to their institutions, and it may be safely assumed that many other medical superintendents do likewise. The 38 per cent of States which comprise the group listed above are merely those that recognize a *fait accompli.* In other words, there is almost general acceptance of the fact that the diagnosis of the feebleminded is based on mental tests and largely a matter for psychologists rather than for physicians who may or may not be psychiatrists, lacking special experience with mental deficiency. It is, however, devoutly to be hoped that other States do not follow the example of Ohio and Oklahoma in deciding the mental test limits which will govern psychologists in their diagnoses.

The setting up of an educational criterion is even less defensible, even though the English practice provides a bad example. The American jurisdictions that base their definition of feeblemindedness directly on educational incapacity are Washington, Texas, and Idaho, the last two in relation to children only, switching to a social definition for adults. Washington admits to its Western State Custodial School residents of the state between the ages of six and 21 years, "who by reason of defective intellect are rendered unable to acquire an education and/or training in the common schools" (Pierce's Washington Code, 1939, 4858-78). It should be noted that the incapacity must arise "by reason of defective intellect," thus presumably making it necessary to employ mental tests to demonstrate the deficiency.

Texas also uses an educational standard but phrases the reason for the incapacity in very peculiar terms. The article is worth quoting in full: "A feebleminded child, as defined herein, is one of such feeble mental or moral powers as to be unable to profit by the ordinary methods of education as employed in the common schools. A feebleminded adult is one who is unable under ordinary circumstances to protect and support himself as a law-abiding citizen because of lack of mental power" (Vernon's Texas Statutes, 1948, Art. 3233). What the ordinary circumstances might be under which a Texan needs mental power to protect himself could give grounds for much legal argument, as would also the recognition of those *morally* unable to acquire an education. The Texas definition is unique and interesting but one that is unlikely to be adopted elsewhere.

Nevertheless, some other States incorporate, along with other criteria, such as incapacity for self-management, some reference to the educational test in their definitions. Idaho admits to its State School and Colony children and youths "who are marked with such mental peculiarities as to deprive them of the school benefits and training provided for youths of normal faculties" (Idaho Code, 1932, Sec. 64-501). Mississippi sets forth, as one of the criteria of feeblemindedness, inability "to profit by ordinary public instruction" (Mississippi Code, 1942, Sec. 6764). North Dakota also admits as feebleminded those "whose defects prevent them from receiving proper training and instruction in the public schools" (N. Dakota Code of 1943, Sec. 25-0404). South Dakota phrases the criterion a little differently, so as to include individuals "who by reason of mental deficiency are incapable of doing the work of the grades in the public schools in a reasonable ratio to their years of life" (South Dakota Code of 1939, Sec. 30-0402).

Utah also uses an educational criterion, but in conjunction with social incapacity. Its law admits to the State institution persons "whose defects prevent them from receiving proper instruction and training in the public schools" (Utah Code, Ann., 1943, Sec. 85-8-16). As an alternative to the need for care in an institution, Wyoming also uses the fact that the alleged feebleminded, if a child of school age, "because of inadequate development, cannot be properly trained in a public school" (Wyoming Compiled Statutes, 145, Sec. 19-417). Montana follows the same rule of combining criteria by admitting "subnormal persons, whose defects prevent them from receiving proper instruction and training in the public schools," along with those incapable of managing themselves and their affairs with ordinary prudence. It is interesting to note that persons "of confirmed immorality" are *excluded* from admission (Chap. 183, Sec. 5, previously cited in an official letter from the Montana State Training School).

We might wonder, at this point, whether the "defects" which prevent these persons from receiving proper instruction and training in the public schools may not be sometimes more attributable to the school systems than to the persons involved.

Kansas, in establishing a State Training School, provided that "the State Department of Social Welfare shall admit into said Training School all idiotic and imbecile children who are incapable of receiving instruction in the common schools." Thus educational incapacity is made the condition for admission to the institution but is not necessarily the basis for definition. In any case, the qualification is unnecessary since idiots and imbeciles are not educable in the common schools. Apparently Kansas makes no statutory provision for a higher grade of feebleminded (General Statutes of Kansas Annotated, 1949, Sec. 76-1605).

Summarizing these data, we may point out that the jurisdictions relying to a greater or less degree on an educational criterion are Idaho, Kansas, Mississippi, Montana, North Dakota, South Dakota, Texas, Utah, Washington, and Wyoming. The English pattern is again traceable throughout the phraseology of their laws. These states number in all 10, or 20 per cent of the total. This is one-third of all jurisdictions with statutory definitions.

Though it is true that various degrees of educational incapacity go with feeblemindedness, we doubt if the citation of this fact strengthens the definition. It by no means follows that all the educationally handicapped are feebleminded. The rolls of famous men are spotted all over with the names of those whom their teachers considered dull. Churchill, Lord Clive of India, and Sir Walter Scott are cases in point. Conversely, there are many who do not fulfil their early scholastic promise, lacking other qualities of initiative or temperament essential to success.

We have already mentioned that the English Act of 1927 sets forth a social criterion to which is added, as a consequence, that the high-grade feebleminded "require care, supervision, and control for their own protection or for the protection of others." Most American jurisdictions which have legal definitions have also combined condition and consequence. It would simplify matters if it were recognized that there are two problems involved, first, who are the feebleminded, and, secondly, what should be done with them when recognized. As we have pointed out, the incorporation of the educational criterion with the social is in recognition of the fact that children have not, in the adult world, yet had the opportunity to demonstrate social competency. But we would submit that the social inadequacy should also, by means of tests of known correlation with the social criterion, be demonstrable at the childish level, in comparison with ordinary children. The real difficulty is not that social incompetence in individuals, whether children or adults, cannot be recognized, but that this inadequacy may arise from causes other than mental deficiency. We have known of instances where social workers have regarded the mere fact of social failure as evinced by delinquency as proof of mental defect, a view which would lead us back to the wholesale diagnoses of *The Kallikak Family*.

Elaine Kinder (6), author of an excellent article in the *Encyclopedia of Psychology*, is careful to emphasize the distinction. She writes:

> The distinction between socially inadequate individuals of low intelligence, and those of the same intelligence level who are capable of "managing themselves and their affairs," and do not require "care, supervision or control for their own protection or for the protection of others" is primarily social and calls attention to the over-all nature of the condition of mental deficiency. "Only limitation in intellectual ability and social competence," she remarks elsewhere, "is common to all defectives."

We have not attempted any comprehensive research into the case law on the subject under discussion, even in the jurisdictions in which no statutory standards exist. However, an annotation entitled "Test or criterion of mental condition within contemplation of statute providing for commitment of persons because of mental condition" found in 158, American Law Reports, Annotated, at page 1220, cites only one case relating to feebleminded persons. This is Re Masters, 216 Minn. 553 (1944), 13 N. W. 2d 487, which is reported in 158 A. L. R. at page 1210. As noted in this report, Minnesota had no statutory definition of feeblemindedness at that time, but by 1949 had adopted one based on the social criterion. In construing the state's commitment statutes the court adopted a "socio-legal" definition of feeblemindedness. For convenience and brevity we quote a portion of the reporter's syllabus, as follows:

> (1). For the purposes of determining whether or not one is feebleminded and subject to commitment and detention as such, the court adopts the definition of "feeblemindedness" proposed by the British Mental Deficiency Committee as follows: "A condition of incomplete development of mind of such a degree or kind as to render the individual incapable of adjusting himself to his social environment in a reasonably efficient and harmonious manner and to necessitate external care, supervision, or control."
>
> (2). Feeblemindedness, viewed from a socio-legal rather than a purely medical standpoint, is not necessarily a permanent and incurable condition.
>
> (3). Psychological definitions phrased in terms of mental age and intelligence quotients are too arbitrary to be applied in cases involving the adjudication, or relief from adjudication, of one as feebleminded and commitment to and retention in an institution for feebleminded persons; hard and fast divisions between groups on the basis of such tests cannot and should not be made.
>
> (4). While psychological tests are convenient tools for indicating mental retardation, test results should ordinarily not be considered sufficient, much less conclusive, except at lower levels, in a proceeding for the adjudication or relief from adjudication of one as feebleminded.

The opinion of the court was delivered by Justice Streissguth and portions of it are so pertinent to the subject of this article that the facts involved are worth summarizing briefly.

A Mrs. Rose Masters had been adjudged feebleminded, and a probate court disallowed a petition for her legal restoration to mental capacity. On appeal, the district court affirmed this decision that the woman was feebleminded, and it was this decision that was reversed by the Supreme Court.

The woman was married in 1923 and had become the mother of 10 apparently normal children. In the depression period the family went on relief, so that a county social agency in 1937 began the supervision of the family. Reports from this agency showed that the woman's housekeeping was very inferior, the meals were inadequate, the children unkempt and unattractive, the house was dirty. Five of the children were finally placed in an orphanage, and both parents adjudged feebleminded and the mother committed to the State institution at Fairbanks.

The secretary of the local welfare board and the county nurse suggested in their report on the woman that the neglect of home duties might have been due to "repeated childbirth which has exhausted her."

Three of the woman's neighbors testified that in their opinion she was "of sound mind." One stated that the children seemed healthy and normal; another, that though Mrs. Masters was a poor housekeeper, she was a good mother. A doctor testified that she was "better than a high-grade moron. . . . She may be a little substandard, but there are a lot of people substandard."

On the stand the woman showed she was able to show up well under lengthy examination, could do simple multiplication, could punctuate and spell well in letter writing, and had a fair vocabulary, and did very satisfactory machine work at the institution. A psychologist who held a master's degree in educational psychology and had eight years' experience in mental testing was accepted as an expert and testified that Mrs. Masters had a mental age of 10 years 4 months, with an intelligence quotient of 64.[7]

In support of the court's opinion Justice Streissguth quoted from C. W. Louttit's textbook, *Clinical Psychology* (8), which discusses the question of feeblemindedness, and cites the British definition which the court adopted, and further asserts that it "seems hardly justifiable to call a man feebleminded because of low test performance of poor school attainment when he is successfully adjusting to the economic and social environment."[8] Louttit also points out that such adjustment may be temporary and not permanent.

Under special circumstances of employment or supervision, a feebleminded adult person may be self-sufficient and self-supporting outside an institution but he is still feebleminded. It is independent self-management and self-support of which the feebleminded adult is incapable, and that is the reason why he requires external care, supervision, and control.

The opinion goes on to state that because temporary social maladjustment does not argue permanent social incapacity, feeblemindedness is therefore not incurable. This difficulty would be avoided if phrasing of the criterion adopted by the judge were, "incapable of *independently* adjusting himself to social environment." As illustrating this point, we have in mind a woman inmate of our feebleminded institution in Hawaii who was released and married. She appears to be adjusted in a reasonably harmonious manner, but several of her children are feebleminded and delinquent; she is a wretched housekeeper, and can assume hardly any of the responsibilities of a housewife.

---

[7] It is evident from this *IQ* that the psychologist was using 16 years as a divisor in calculating the intelligence quotient, a practice which has been discarded by competent clinical psychologists. Using 14 years as the divisor (representing the average mental age of adults), the woman's *IQ* would have been 73, above the formerly accepted *IQ* level of feeblemindedness.

[8] The first four years of Dr. Louttit's experience in clinical psychology were spent as assistant to the senior author of this article. This may in part account for the similarity of Judge Streissguth's position in this matter and our own.

But, as a matter of fact, the harmony of adjustment is on the part of a husband and neighbors who put up with her presence in the home and community. Take away these adventitious props and she would at once need institutional care. She is not, however, actively delinquent and the people who know her, including the welfare worker, are very tolerant. Nevertheless, this kind of social adjustment could hardly be considered proof of adequate mental capacity.

The laws relating to the feebleminded in the majority of American jurisdictions do depend on this social criterion. So many of them follow along similar lines that it is hardly necessary to quote them all. California provides a typical example:

> As used in this code, "mentally deficient persons" means those persons, not psychotic, who are so mentally retarded from infancy or before reaching maturity that they are incapable of managing themselves and their affairs independently with ordinary prudence, or of being taught to do so, and who require supervision, control and care for their own welfare or for the welfare of others, or for the welfare of the community. (California Welfare and Institution Code, 1951, Section 5250, as amended).

From our discussion of the problem, it may be judged that this California law would be regarded as one of the most comprehensive. It excludes the psychotic, it refers to early onset or appearance, it emphasizes the mental retardation which presumably must be present, it states the social inadequacy, the need for institutional treatment, and in addition sets forth the incurability of the condition. Apparently the only thing California left out is reference to the educational incapacity.

Louisiana defines a mental defective as "a person who is not mentally ill but whose development is so retarded that he has not acquired enough self-control, judgment, and discretion to manage himself and his affairs, and for whose own welfare or that of others, care, supervision, guidance, or control are necessary or advisable. The term includes feebleminded, idiot, and imbecile" (Louisiana R. S. 1950, 28:2).

Mississippi also has aimed at comprehensiveness, with the result that its statute is unwieldy. It includes as criteria mental inferiority existing from an early age, inability to care for themselves, to profit by ordinary public school instruction, to compete on equal terms with others, or to manage themselves or their affairs with ordinary prudence, and who consequently constitute menaces to the happiness or safety of themselves and of other persons in the community and require care, supervision, and control either for their own protection or the protection of others. Lastly the statute recites the divisions of feebleminded into idiots, imbeciles, and "morons or high-grade feebleminded persons." The only unmentioned features are the suggested psychometric boundaries. Such a prolix and repetitive law may be in good legal tradition, but it is doubtful whether the statute is made plainer, or more applicable, thereby (Mississippi Code, 1942, 6764, Title 24, Chap. 6).

Nebraska is the only State that introduces an industrial criterion along with the social—in our opinion, a worthwhile addition. In its statute, the mental defectiveness must be so pronounced that the person is incapable of managing himself and his affairs, and of subsisting by his own efforts, or of being taught to do so (Revised Laws of Nebraska, 1950, Chapter 83-219).

In general, these are the judisdictions that in one form or another set forth the social criterion: Alabama, Arizona, California, Colorado, Delaware (partially), Hawaii, Georgia, Idaho, Illinois, Iowa, Kentucky, Louisiana, Minnesota, Mississippi, Montana, Nebraska, New Mexico, New York, North Dakota, Ohio, Pennsylvania, South Carolina, Tennessee, South Dakota, Texas, Utah, Virginia, West Virginia. Thus the total number which recognize the social basis is 28, or 56 per cent of the total. This constitutes over 80 per cent of the States which have some statutory definition. Delaware is included because, though it has no legal definition, it provides that an individual may be committed when it appears to the judge that it would be detrimental to the community for such a person to remain at large (Revised Code of Delaware, 1935, Chapter 70, Sec. 113). Altogether, 24 states mention the need for institutional care.

An advantage that would result in separating this provision from the definition stems from the fact that well-trained defectives could be released from the institution under conditions of only limited supervision without casting doubts on the validity of the original diagnosis. Obviously, if the law specifically states that to require care, supervision, and control is part of the proof of feeblemindedness, then dismissal or parole from the institution would indicate they are no longer feebleminded. Very young children, of every degree of mental development, also require care, supervision, and control, and so these conditions should not be part of the definition.

Except for the widespread confusion and uncertainty, it might seem presumptuous to present any form of a model statute, but under the circumstances we shall tentatively set forth five provisions which we consider such a statute should include:

1. *Definition.* Feebleminded persons are those who by reason of per-

manently retarded or arrested mental development existing from an early age are incapable of independent self-management and self-support.

2. *Commitment.* A feebleminded individual needing care, control, or supervision should be committed to a state institution only by a court of record on the recommendation of an examining board.

3. *Examining Board.* This should include a physician or psychiatrist, and a clinical psychologist, each of whom is qualified by special experience in the diagnosis of the feebleminded. The board or commission would be strengthened by the addition of a lay person, preferably a lawyer who would act as chairman.

4. The alleged feebleminded person, and his closest relatives, should be given due notice of the proceeding by which the adjudication is sought, and have the right to be represented by counsel. The examining board would be authorized to act only by virtue of a reference from the court to which an application has been made.

5. *Certification.* The board should (1) determine the degree of mental arrest or retardation by the use of psychological tests of proved validity for the diagnosis of feeblemindedness; and (2) should certify, setting forth findings based on recorded evidence, that the alleged feebleminded person is incapable of self-management and self-support.

6. *Discharge.* This should be based on the recommendation of the examining board and the superintendent of the institution on the grounds that the individual is not likely to be a burden on the community. The individual should be entitled, as a matter of right, to a review of his case at reasonable intervals, upon his own request or the request of his close relatives.

Briefly stated, the above, as a basis of statutory definition, has the following advantages. It is clear and concise, yet of adequate scope. The definition covers both children and adults. It implies that the condition is permanent and it combines the social, psychological, and industrial criteria. It excludes the psychotic, the normally immature, and those socially inadequate for other reasons than mental deficiency.

The outline has the further advantage of separating definition from disposition, but most of all, it provides maximum protection to the alleged feebleminded, by putting diagnosis in the hands of fully qualified examiners. It takes legal cognizance of the fact that psychological tests are universally used for diagnostic purposes, but abstains from setting up psychometric standards that may be subject to change. We believe that a statute based upon the above considerations would have the characteristics of a good law—

clarity, conciseness, and comprehensiveness, with proper provision in respect to the rights of the individual.

REFERENCES

1. BALLER, W. R. Study of the present social status of a group of adults, who, when they were in elementary school, were classified as mentally deficient. *Genet. Psychol. Monog.*, 1936, **18**, 167-244.
2. BURT, C. The Subnormal Mind. London: Oxford Univ. Press, 1935.
3. DOLL, E. A. Notes on the concept of mental deficiency. *Amer. J. Psychol.*, 1941, **54**, 116-124.
4. GODDARD, H. H. The Kallikak Family. New York: Macmillan, 1912.
5. ———. Feeblemindedness: A question of definition. *Proc. Amer. Assn. Study Feebleminded.* June, 1928.
6. KINDER, E. Feeblemindedness. In the *Encyclopedia of Psychology.* New York: Philosophical Library, 1946.
7. LEWIS, N. D. C. Outlines for Psychiatric Examinations. New York State Dpt. Mental Hygiene, 1943.
8. LOUTTIT, C. M. Clinical Psychology. New York: Harper, 1936.
9. TERMAN, L. M. The Measurement of Intelligence. Boston: Houghton Mifflin, 1916.
10. TIZARD, J. The abilities of adolescent and adult high-grade male defectives. *J. Ment. Sci.*, 1950, 1-19.
11. WEIHOFEN, H. Hospitalizing the mentally ill. *Michigan Law Review,* 1952, **50**, 837-872.

Juvenile Court
Honolulu 9, Hawaii

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 107-117.

PERFORMANCE ON A FOLLOWING TRACKING TASK (MODIFIED SAM TWO-HAND COÖRDINATION TEST) AS A FUNCTION OF THE PLANES OF OPERATION OF THE CONTROLS*[1]

Department of Psychology, University of Rochester

EUGENIA B. NORRIS AND S. D. S. SPRAGG

A. INTRODUCTION

In many types of machine control operations, operators are required to move the controls more or less continuously in response to some type of visual indicator, making what Brown and Jenkins (1) have called continuous movement reactions. So far, relatively little detailed knowledge is available concerning the factors which affect the accuracy with which such movements can be made. One of the questions that arises is that of the optimum relationship between the operator and the controls, and the controls and the display or indicator. Some studies have been made (2, 3, 4, 5, 7, 9, 10, 11) of these relationships. In general, simple one-handed tasks have been used in these studies; in those in which a two-handed task was performed the responses of the two hands were fairly discrete. Some of these studies have indicated that the optimum relationships between the controls and the indicator are (*a*) those in which the control moves the indicator in the "expected" direction, e.g., moving the control up moves the indicator up, and (*b*) those in which there is continuity between the axis of movement of the control and that of the indicator.

This report is the first of a series of studies investigating the optimum relationships between operators and controls, and controls and display, using a *complex two-hand* coördination task. The present report is an investigation of the performance of such a task as a function of the planes of rotation of the controls. Subsequent studies will examine performance as a function of the direction of rotation of the controls in relation to the direction of movement of the display, the optimum ratios between the amount of movement of the control and the amount of movement of the display, etc.

---

*Received in the Editorial Office on August 27, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was conducted as part of a research program on human motor performance carried out under Contract No. N6 onr-241, T. O. VI between the University of Rochester and the Special Devices Center, Office of Naval Research, to whom these results have been reported in mimeographed form as Technical Report No. SDC 241-6-3.

B. Subjects

The subjects were 100 male undergraduates at the University of Rochester. Their ages ranged from 18 to 34 years and both right and left handed subjects were used. They were paid for their services. Some of the subjects had been tested on the SAM Two-Hand Coördination Test as part of the selection battery at Air Forces Classification centers during the war. They were, however, well distributed among the various experimental conditions. To randomize the groupings the subjects were assigned to the various experimental groups in the order of their appearance.

C. Apparatus

The apparatus used was a modification of the SAM Two-Hand Coördination Test (CM101B). A complete description has been given by Melton (6, Ch. 5). This apparatus is designed to measure the ability to coordinate the movement of a target follower, by turning two cranks, in response to a visually perceived target moving at varying rates along an irregular pathway. The modification consisted of mounting the control cranks on pivoting stands which made it possible to place each control in any one of several possible planes of rotation.

A reversible coupled synchro drive system was used. There were two pairs of Type 5, 110 v, 60-cycle synchros, the transmitter unit being mounted on the pivoted stand with the crank connected to it and the receiver unit connected to the driving mechanism of the Two-Hand Coördinator. By means of double pole, double throw switches, the fields in each of the transmitter synchros could be reversed, thereby reversing the direction of movement produced by the rotation of the cranks.

Figure 1 shows the essential features of the apparatus from the subject's position. The subject's task was to keep the target follower (F) any place on top of the target (E), controlling the target follower by means of the two cranks (A and B). The entire turntable (G) rotated in a clockwise direction, the target (E) moving with it but also moving in an irregular manner in the curved slot (H). One trial consisted of one complete revolution of the target, taking one minute. One cycle consisted of four one-minute trials, the pattern followed by the target for each of the four trials being different. There was a 15-second rest period between each trial and a warning buzzer sounded approximately one second before the beginning of a trial. The controlling mechanism was automatic and the scoring was by means of a Standard Electric Timer calibrated in thousandths of a minute.

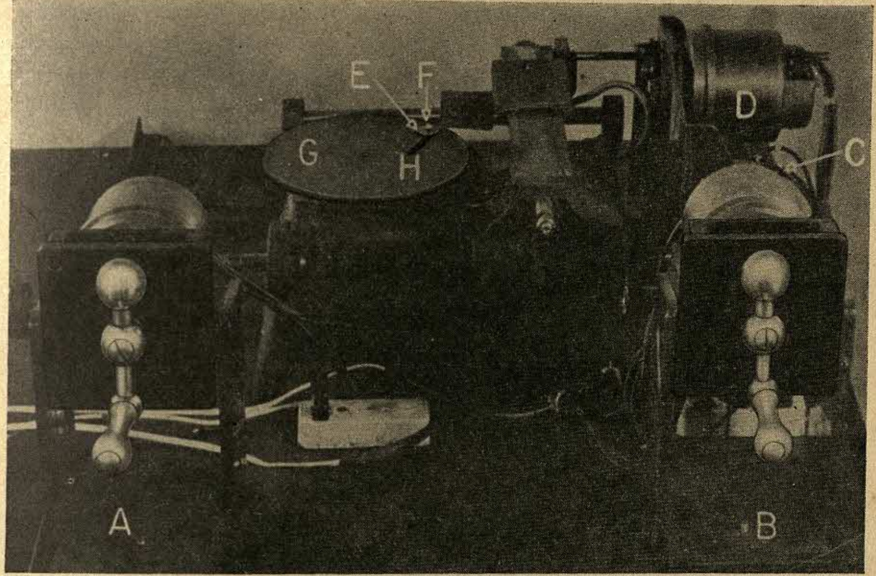


FIGURE 1
THE MODIFIED TWO-HAND COÖRDINATOR

The target follower (F) was a small button mounted at the end of the leaf of a microswitch. When the button was on the target it closed the microswitch, activating the circuit to the clock.

On the preliminary trials the cranks were separated by a distance of 9 inches and on the experimental trials the separation was kept constant at 19 inches for all the groups except the control group. The effective width of the target was .8750 inches; 3.86 revolutions of a crank were necessary to move the target follower across the target, or each revolution of the crank moved the target follower .2267 inches. The diameter of the arc produced by the cranks was four inches.

D. PROCEDURE

Ten groups of subjects (consisting of 10 *S*s each) were used, the only difference between conditions for the several groups being the planes of rotation of the cranks for the experimental trials. Each *S* was tested at two sessions, preliminary and experimental, with five to seven days intervening between them. At the first session, which lasted approximately 10 minutes, *S* was given instructions for the test (which followed very closely those reported by Melton (6, pp. 192-193) for the Two-Hand Coördination

Test), then was given four trials with the control cranks placed so that they rotated in the vertical plane perpendicular to *S*'s frontal body plane and facing each other. Counterclockwise rotation of the right crank moved the target follower away from the *S*, and counterclockwise movement of the left crank moved the target follower to the right. These preliminary trials were given in order to obtain an estimate of the initial equality of the groups on this task.

In the second or experimental session the *S*s had their attention called to the fact that the control cranks were now in a new combination of positions (except for Group 10, for whom the original position was continued), and the essential parts of the general instructions were repeated. Then each *S* was given eight trials, with the cranks placed in one of the position combinations described below.

At this second session the apparatus was arranged so that counterclockwise rotation of the right crank moved the target follower toward the *S*, and of the left crank, moved it to the left. This change in the direction of movement of the target follower between the preliminary and experimental sessions was made in order to retard the rate of learning. If many *S*s made a perfect score, an artificial upper limit would have been placed on the distribution of scores. That this technique had some success is seen from the fact that only two subjects ever made perfect scores. One subject in Group 1 made a perfect score on two trials and another, in Group 4, on one trial.

In the experimental groups, three planes of crank rotation were used from among the possible planes. The combination of planes for the two cranks for each of the groups is presented in Table 1. In this table the letters *A*, *B*, and *C* refer to the planes of rotation of the cranks; *A* is vertical and parallel to the frontal body plane, *B* is vertical and perpendicular to the frontal body plane, and *C* is horizontal. Each group had a combination of these planes as indicated in the table.

It may be seen that for Groups 1, 2, 3, and 10 both hands were operating

TABLE 1
COMBINATION OF PLANES OF CRANK ROTATION FOR EACH GROUP

Group	Left hand	Right hand	Group	Left hand	Right hand
1	A	A	6	C	A
2	B	B	7	A	C
3	C	C	8	B	C
4	A	B	9	C	B
5	B	A	10	Same as preliminary trials.	

in the same plane of rotation, while for all other groups the two hands operated in different planes of rotation.

Group 10 may be called a control group in the sense that for their experimental trials the cranks remained in the same position as for the preliminary trials, even though for this group, as for all the groups, the control-display movement relationships were changed between the preliminary and the experimental trials. Group 10 was run in order to have some basis for ascertaining whether the increase in the level of performance was a function entirely of learning or whether it might reasonably be assumed to be a function of the changes in the planes of rotation of the cranks between the preliminary and experimental sessions.

E. RESULTS

The results are expressed in terms of the mean time on target for the four preliminary trials and the eight experimental trials for each subject.

1. *Preliminary Trials*

Table 2 is a summary table of the analysis of variance of the preliminary trials. Since the resulting F ratio is less than one, it is reasonable to assume that, in terms of initial ability, the differences between the groups are no greater than would be expected by chance.

Table 3 gives the mean time on target and the standard deviation for each of the 10 groups for the four preliminary trials. The maximum difference between the means of any two of the groups was 77.0. A differ-

TABLE 2

SUMMARY TABLE FOR THE ANALYSIS OF VARIANCE OF THE PRELIMINARY TRIALS

Source	*df*	Sums of squares	Mean square	*F*
Between groups	9	74,843.5	8,315.9	<1
Within groups	90	1,356,998.5	15,077.7	
Total	99	1,431,842.0		

TABLE 3

MEAN TIME ON TARGET AND STANDARD DEVIATION FOR THE PRELIMINARY TRIALS (GROUPS PRESENTED IN RANK ORDER)

	Groups									
	1	7	4	10	3	6	9	2	5	8
Mean time on target	455.2	454.6	438.8	434.1	425.9	420.1	416.7	384.4	382.6	378.2
SD	117.7	76.4	114.2	138.3	150.8	105.2	84.5	131.6	121.8	105.4

ence of 111.9 would be necessary for significance at the five per cent level of confidence.

2. *Experimental Trials*

An analysis of variance for the experimental trials, Table 4, indicates that the obtained differences between the means of the groups would be found by chance less than one time in a thousand ($F = 4.181$; $df = 9$ and 90); it can thus be concluded that the planes of rotation of the handles did make a significant difference in the performance of this task.

TABLE 4
SUMMARY TABLE FOR THE ANALYSIS OF VARIANCE OF THE EXPERIMENTAL TRIALS

Source	*df*	Sums of squares	Mean square	*F*
Between groups	9	505,456.5	56,161.8	4.181*
Within groups	90	1,208,669.8	13,429.6	
Total	99	1,714,126.3		

*Significant at the .001 level.

TABLE 5
MEAN AND STANDARD DEVIATION OF THE TIME ON TARGET FOR THE EXPERIMENTAL TRIALS (GROUPS PRESENTED IN RANK ORDER)

	Groups									
	1	4	7	2	6	8	9	3	5	10
Mean time on target	759.5	746.7	647.9	643.7	631.9	619.1	609.5	582.2	562.2	519.7
SD	131.3	116.1	96.7	131.0	65.4	79.5	139.6	122.1	122.2	126.7

Table 5 presents the mean time on target and its standard deviation for the experimental trials for each of the groups. Figure 2 shows the same data graphically. Calculating the *t*'s, using the "within groups" variance of the analysis of variance as the error estimate, it was found that a difference of 137.7 between the means of any two of the groups was significant at the one per cent level of confidence, one of 103.6 at the 5 per cent level, and one of 98.4 at the 6 per cent level.

Table 6 presents the significance level of the differences between Groups 1, 4, and 10 and the other groups.

TABLE 6
SIGNIFICANCE LEVELS OF DIFFERENCES BETWEEN THE VARIOUS EXPERIMENTAL GROUPS

Group		Groups	Level of confidence
1	>	8, 9, 3, 5, 10	1%
1	>	7, 2, 6	5%
4	>	3, 5, 10	1%
4	>	2, 6, 8, 9	5%
4	>	7	6%
10	<	1,4	1%
10	<	7,2, 6	5%

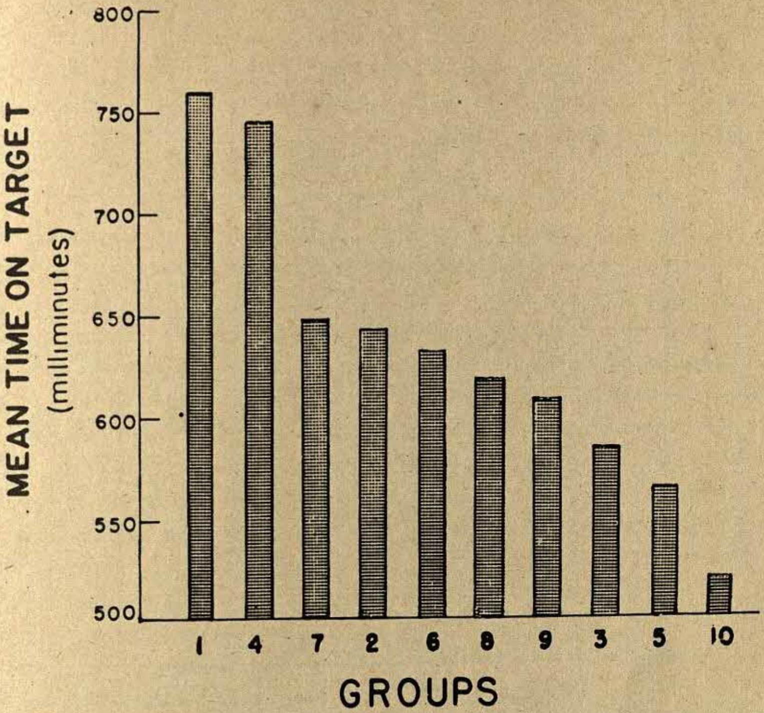


FIGURE 2
Mean Time on Target for Each Group (Experimental Trials)

It will be noted that Groups 1 and 4 are superior to all other groups at the 6 per cent level of confidence or better, while Group 10 is significantly poorer than Groups 1, 4, 7, 2, and 6 at the 5 per cent level or better.

Figure 3 presents the learning curves for each of the groups in terms of the mean time on the target for each of the eight experimental trials. It will be noted that on each of the trials Groups 1 and 4 are superior to all the other groups and, with the exception of Trials 7 and 8, Group 10 is inferior to all other groups. However it is of some interest to note that, although the over-all performances of the groups show considerable differences, the slopes of the learning curves are approximately the same, indicating that rate of learning was essentially similar for all the groups.

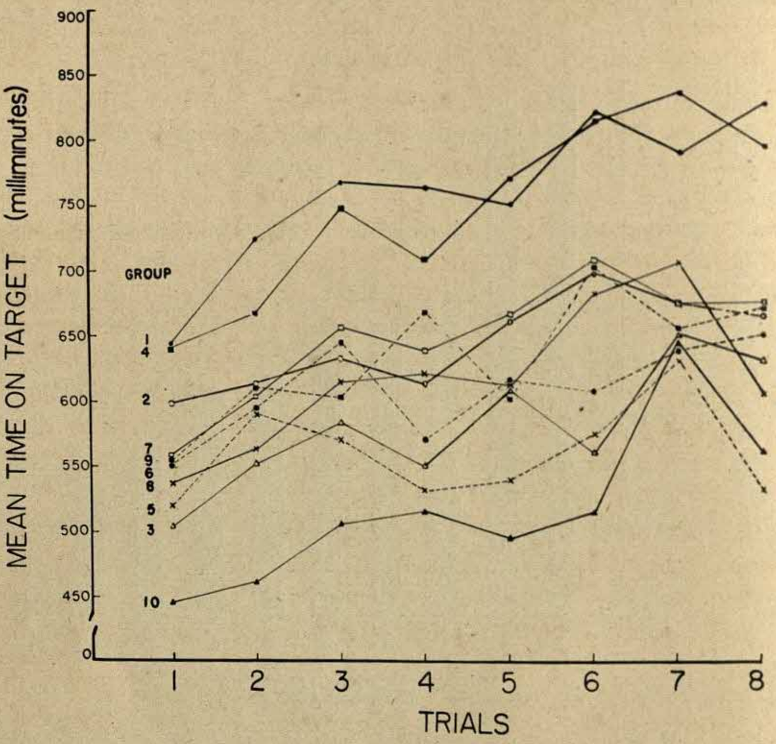


FIGURE 3
Learning Curves for Each Group (Experimental Trials)

F. Discussion

The results of the analysis of the preliminary trials showed that the 10 groups could be considered equivalent before they were given the various experimental conditions. Therefore, differences between performances of the groups on the experimental sessions can reasonably be described as a function of the experimental conditions imposed on each group. Analysis of the results indicates that performance of this complex eye-hand coördination task varies significantly as a function of the planes of rotation of the controls. In terms of the *t* tests and also in terms of the learning curves, it is reasonable to assume that Conditions 1 and 4 result in better performance of the task than the other conditions tested. For Group 1, it will be remembered, both cranks rotated in the vertical plane parallel to the frontal body plane. In this condition the cranks are in a position with which many of the *S*s have probably had some previous experience, since machine controls are frequently set up to be operated in this manner. Also it should be noted that with the conditions of Group 1 the subject turns both arms through a cone with the result that the effective mass to be accelerated is less than in any of the other conditions, in which at least one arm must move forward and backward like a piston.

For Group 4 the cranks were placed as follows: the right crank rotated in the vertical plane perpendicular to the frontal body plane and the left crank in the vertical plane parallel to the frontal body plane. Therefore, for Group 4, rotation of the right crank in a plane perpendicular to the frontal body plane caused the target follower to move in the same plane, and rotation of the left crank in a plane parallel to the frontal body plane caused the target follower to move in the same plane. There was, therefore, *continuity* between the movement of the control and the target follower. It might be said that manipulation of the controls produced the "expected" movement of the target follower. The efficacy of a relationship in which there is continuity between control movement and display movement seems to have some generality as indicated by the results of certain other studies in simpler task situations (2, 5, 9, 10).

It is not surprising that the control group (Group 10) had the lowest rank, since for this group the cranks rotated in the vertical plane perpendicular to the frontal body plane and facing each other. This made it necessary to grasp the cranks "backward," that is with little finger on the inner side of the crank, a relatively awkward position. It is interesting to note, however, that only Groups 1 and 4 were significantly better than Group 10 at the one per cent confidence level, with Groups 7, 2, and 6 being better than Group 10 at the 5 per cent level. Taking the mean of Group 10 as the score to be expected merely on the basis of learning over the eight trials, the results for Groups 8, 9, 3, and 5 may be interpreted as a function of learning only, with no significant contribution to performance contributed by position.

It was thought that the groups in which both of the controls were in the same plane (Groups 1, 2, and 3) might show some consistent difference from the other conditions. However, the results show that this was not the case, the ranks for these groups being as follows: Group 1, Rank 1; Group 2, Rank 4; and Group 3, Rank 8.

G. SUMMARY

The present study was designed to investigate performance of a two-hand coördination task as a function of the planes of rotation of the controls. The SAM Two-Hand Coördination Test was modified so that each crank control could be placed so that it rotated in any one of a variety of planes. The three planes used were: (*a*) horizontal, (*b*) vertical and parallel to the frontal body plane, and (*c*) vertical and perpendicular to the frontal body plane. The nine resulting possible combinations of these planes necessitated nine experimental groups plus a control group. Each group (10 subjects) was given four preliminary trials to demonstrate the equality of the groups. On these trials the controls were in the same plane of rotation for all 10 groups of subjects. At a second session, eight experimental trials were given, each of the nine experimental groups being assigned to one of the combinations of planes tested and the control group having the same combination of planes as in the preliminary trials.

The results were analyzed in terms of the mean time on target for each subject for the preliminary and for the experimental trials. There were no significant differences between the groups on the preliminary trials. For the experimental trials the following two conditions were found to be superior to all others: (*a*) with the two cranks rotating in the vertical plane parallel to the frontal body plane (Group 1), and (*b*) with the right crank rotating in the vertical plane perpendicular to the frontal body plane and with the left crank rotating in the vertical plane parallel to the frontal body plane (Group 4).

REFERENCES

1. BROWN, J. S., & JENKINS, W. O. An analysis of human motor abilities related to the design of equipment and a suggested program of research. (Ch. 3 in Fitts, P. M. (*Ed.*), *Psychological Research on Equipment Design.*) Report No. 19, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.
2. CARTER, L. F., & MURRAY, N. L. A study of the most effective relationships between selected control and indicator movements. (Ch. 10 in Fitts, P. M. (*Ed.*), *Psychological Research on Equipment Design.*) Report No. 19, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.
3. CRAIK, K. J. W., & VINCE, M. A. Psychological and physiological aspects of control mechanisms with special reference to tank gunnery: I. Military Personnel Research Committee, Medical Research Council, Great Britain BPC 43/254, August, 1943.
4. ———. Psychological and physiological aspects of control mechanisms with special reference to tank gunnery: II. Military Personnel Research Committee, Medical Research Council, Great Britain BPC 44/322, March, 1944.
5. GRETHER, W. F. Direction of control in relation to movement in one dimensional tracking. AAFH AMC Eng. Div., TSEAA 8 WFG mec., October, 1947.

6. MELTON, A. W. (*Ed.*). Apparatus Tests. Report No. 4, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.

7. MITCHELL, M. J. H. Direction of movement of machine controls: III. A two-handed task in a discontinuous operation. Ministry of Supply, Psychological Laboratory, Cambridge, August, 1947.

8. ————. Direction of movement of machine controls: V. A two-handed performance in a continuous task. Applied Psychology Panel, Psychological Research Unit, Psychological Laboratory, Cambridge, May, 1949.

9. REED, J. D. Factors influencing rotary performance. *J. of Psychol.,* 1949, **28**, 65-92.

10. VINCE, M. A. Direction of movement of machine controls: I. Flying Personnel Research Committee, Ministry of Supply, No. 637, August, 1944.

11. VINCE, M. A., & MITCHELL, M. J. H. Directions of movement of machine controls: II. Ministry of Supply, Psychological Laboratory, Cambridge, November, 1946 (FPRC No. 637, cont.).

12. WARRICK, M. J. Direction of movement in the use of control knobs to position visual indicators. (Ch. 9 in Fitts, P. M. (*Ed.*), *Psychological Research on Equipment Design.*) Report No. 19, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.

Department of Psychology
University of Rochester
Rochester 3, New York

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 119-129.

PERFORMANCE ON A FOLLOWING TRACKING TASK (MODIFIED SAM TWO-HAND COÖRDINATION TEST) AS A FUNCTION OF THE RELATIONS BETWEEN DIRECTION OF ROTATION OF CONTROLS AND DIRECTION OF MOVEMENT OF DISPLAY*[1]

Department of Psychology, University of Rochester

EUGENIA B. NORRIS AND S. D. S. SPRAGG

A. INTRODUCTION

The present study is the second in a series investigating the optimum relationships between operators and controls, and controls and displays in a complex coördination task (SAM Two-Hand Coördination Test). The variable investigated is the direction of rotation of controls in relation to the direction of movement of the display. There are numerous situations in which the operator's task is to position or to keep positioned some aspect or element of a display. Frequently he is required to do this by manipulating rotary controls such as knobs or cranks or handwheels. Two sets of general statements may be made about the possible display-control movement relationships, depending upon whether the display movement resulting from movement of a control is perceived as curvilinear or rectilinear. If it is curvilinear then clockwise movement of a control may produce (*a*) clockwise movement of the controlled aspect of the display, or (*b*) it may produce counter-clockwise movement of it. If the display movement is described in terms of a system of rectilinear coördinates, then clockwise rotation of a single control may result in a display movement in one dimension which can be described as up or down, left or right, or toward or away, depending upon the relation of that movement to the operator's orientation.

In many instances, however, display movement is two-dimensional movement in a given plane and two rotary controls are used, one for each coordinate of movement. In such cases clockwise movement of one control

---

*Received in the Editorial Office on August 27, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was conducted as part of a research program on human motor performance carried out under Contract No. N6 onr-241, T. O. 6 between the University of Rochester and the Special Devices Center, Office of Naval Research, to whom these results have been reported in mimeographed form as Technical Report No. SDC-241-6-5.

may produce (*a*) display movement toward the operator's right or (*b*) display movement toward his left. Clockwise movement of the other control may produce display movement (*a*) toward the operator or (*b*) away from the operator. (This would be the case where display movements are in the horizontal plane. If they are in a vertical plane parallel to the operator's frontal body plane—a very common position—then movement of the second control would produce display movement in an upward or downward direction with reference to the operator.)

Previous investigations of such relationships in relatively simple situations (1, 5, 6, 8, 10, 11, and 12) have shown that when the relationship is the "natural" or "expected" one, e.g., clockwise movement of the control producing clockwise movement of the display, performance is significantly better than when this is not the case. However, in these studies the tasks employed were either: (*a*) simple, discrete one-hand settings; (*b*) two-hand tasks with discontinuous movements; or (*c*) simple continuous movements in which the speed of movement was the only factor considered. None of the studies cited has dealt with a following tracking task such as the present one in which a complex continuous two-hand coördination is the task. It is well known that relationships which hold under simple conditions are frequently altered when the conditions are made more complicated or difficult.

In a previous study (7) it was found that placing the control cranks: (I) so that each rotated in a vertical plane parallel to the frontal body plane of the subject; or (II) the left control as in I and the right control rotating in a vertical plane perpendicular to the frontal body plane, resulted in better performance of this complex eye-hand coördination task than any of the other conditions tested. The data revealed no significant difference in performance between these two conditions however, so a decision could not be made at that point as to which was the superior control arrangement.

B. Statement of the Problem

The present investigation was so designed that estimates could be made of the accuracy of performance on the complex two-hand coördination task as a function of: (*a*) the direction of rotation of the controls in relation to the direction of movement of the display; and (*b*) the plane of rotation of the controls relative to the direction of display movement. Four different combinations of direction of rotation and two different planes of rotation were used.

C. SUBJECTS

The subjects were 96 male undergraduates at the University of Rochester. Their ages ranged between 17 and 33 years and both right- and left-handed subjects were used. They were assigned to the various experimental conditions in the order of their appearance and were paid for their services. A few of the subjects had had a brief experience on the Two-Hand Coordination Test several years previously as part of the AAF cadet selection procedure during the last war, but these subjects were well distributed throughout the various experimental conditions.

D. APPARATUS

The apparatus used was a modified SAM Two-Hand Coördination Test, (CM 101B) which has been described in detail in a preceding study (7). The modifications made it possible to reverse the direction of movement produced by rotation of the cranks and to place them in horizontal or vertical planes. The subject's task was to keep a target follower any place on top of a small circular brass target which rotated in a clockwise direction and also moved irregularly in a curved slot. The left crank moved the target follower from left to right and the right crank moved it towards and away from the subject. Each trial was one minute in duration and consisted of one complete revolution of the target. Scoring was by means of a Standard Electric Timer calibrated in thousandths of a minute.

E. PROCEDURE

Two groups of 48 subjects each were used in this experiment, the only experimental difference between the groups being the planes of rotation of the cranks. For Group I both of the cranks were placed so that they rotated in a vertical plane, parallel to the frontal body plane of the subject; for Group II the left crank was placed as for Group I, and the right crank was placed so that it rotated in a vertical plane, perpendicular to the frontal body plane.

The experimental procedure for the two groups was the same. Each subject was tested at one session lasting approximately 20 minutes. The general instructions were the same as in the previous study (7). At the end of the first and each subsequent trial the following instructions were given:

> Get on the target. For the next trial the directions of rotation of one or both handles will be changed. For example, perhaps on the last trial clockwise rotation of the right handle moved the button towards

> you. On this trial, then, clockwise rotation *may* move it away from you. In just a moment you will be allowed to practice with the new condition. Are there any questions? You may now practice. (*After one minute*) Release the handles. Again when you hear the buzzer grasp the handles and when the target moves keep the button on it.

Each subject was given a total of four trials, one on each of the conditions (*A, B, C,* and *D*) listed in Table 1. This table shows, for each experimental condition, the effect of clockwise rotation of each control crank upon movement of the display member (target follower). Counterclockwise movement of the crank will of course produce the opposite effect in each case.

TABLE 1
EXPERIMENTAL CONDITIONS

Experimental condition	Crank	Clockwise rotation of the control crank produces display movement
A.	Left	To the left
	Right	Toward the subject
B.	Left	To the right
	Right	Away from the subject
C.	Left	To the right
	Right	Toward the subject
D.	Left	To the left
	Right	Away from the subject

Since it has been shown that learning of this task is fairly rapid (4), a 4 x 4 latin square design was used in order to control sequence effects. Since there were 48 subjects in each of the two groups, there were 12 replications of the latin square for each group.

There was a two-minute rest period between each trial. During this period the experimenter changed the direction of rotation of the controls and the subject practiced with the new condition for one minute.

From the results of previous investigations it is reasonable to assume that the "expected" or "natural" relationships between the movement of the controls and the movement of the display are those of Condition *B*, that is, the condition in which clockwise rotation of the left crank produces movement to the right, and of the right crank produces movement away from the subject. Under Condition *A* both hands would be operating under "unnatural" conditions, while for *C* and *D* "natural" conditions would prevail for one hand only, the other being under "unnatural" conditions.

The experimental design also makes possible the detection and measurement of interference effects. Since each of the subjects was tested under each of the possible combinations of directions of rotation it was at times necessary for them to make responses to the same stimulus situation under conditions in which the S-R relationships were either completely or partially reversed. It is probable that interference will be manifested in such a situation. One question to be examined in the present study is whether such interference will affect performance differentially depending on whether there is complete continuity between the planes of movement of the control and the display (i.e., in Group II), or whether there is only partial continuity (Group I). It is also of interest in this connection to note that for two sequences of the 4 x 4 latin square used there are two cases of complete reversal; that is, on two successive trials the direction of rotation to produce a given display movement is reversed for *both* hands. In the other two sequences of the latin square no such *complete* reversals occur.

F. Results

Analysis of the data has been directed towards determination of: (*a*) the best combination of directions of rotation of the controls; (*b*) the most effective combination of planes of rotation for the cranks under the present conditions; and (*c*) the relative effectiveness of direction of rotation with the two combinations of planes used.

Figure 1 presents the results graphically. The mean time on target for the 48 subjects for each of the combinations of directions of rotation is presented for each of the two groups.

Results of the analysis of variance for the data of the two groups separately and for Group I versus Group II are presented in Table 2. The error term in this analysis is a pure error-estimate of variation which is independent of the effects of directions, sequence of directions, ordinal position of the direction in a sequence, and the individual differences among the subjects within a sequence. The procedure for obtaining this error estimate has been described by Grant (2). The error term is combined with the square uniqueness term since a preliminary analysis showed that the F ratio for square uniqueness over the error term was not significant at the 5 per cent level of confidence. Directions and the ordinal position of a direction in the sequence are significant at the one per cent level for both groups when tested against the error estimate. An F ratio of the mean square for sequence of directions over the mean square for individual subjects within sequence is not significant at the 5 per cent level. Thus, the performance of the task was found

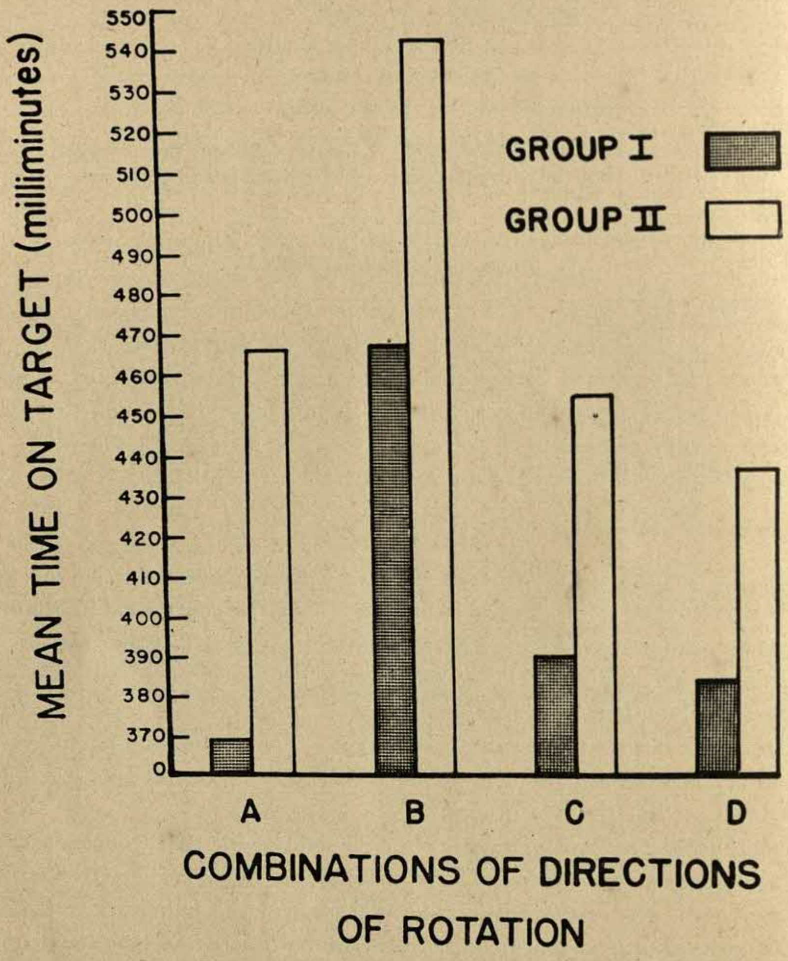


FIGURE 1
SHOWING THE MEAN PERFORMANCE OF EACH EXPERIMENTAL GROUP ON THE FOUR COMBINATIONS OF DIRECTION OF CRANK ROTATION

to be significantly a function of the direction of rotation of the controls, and of the ordinal position of that direction, but not of the sequence of directions used.

Further analysis showed that the combination of directions in Condition *B* (that is, both hands performing under the "natural" conditions) led to better performance of the task for both groups than did any of the other combinations. All the differences were significant at the one per cent

TABLE 2
ANALYSIS OF VARIANCE FOR GROUPS I AND II AND FOR GROUP I VERSUS GROUP II

Group I Source	*df*	*SS*	*MS*	*F* ratio	
1. Directions	3	286,418	95,472	1/5	7.86*
2. Sequence of directions	3	310,510	103,503	2/4	1.88
3. Ordinal position of direction	3	551,387	183,794	3/5	15.14*
4. Individual subjects within sequence	44	2,421,522	55,034	4/5	4.53*
5. Error (Combined term: error and square uniqueness)	138	1,675,249	12,139		
6. Total	191	5,245,086			
Group II Source	*df*	*SS*	*MS*	*F* ratio	
1. Directions	3	308,120	102,706	1/5	7.18*
2. Sequence of directions	3	168,501	56,167	2/4	<1
3. Ordinal position of direction	3	832,411	277,470	3/5	19.40*
4. Individual subjects within sequence	44	2,701,789	61,404	4/5	4.29*
5. Error (Combined term: error and square uniqueness).	138	1,973,792	14,302		
6. Total	191	5,984,613			
Group I versus Group II Source	*df*	*SS*	*MS*	*F* ratio	
1. Groups	1	501,632	501,632	1/2	8.61*
2. Individual subjects	88	5,123,311	58,219	2/3	4.40*
3. Error (Combined term: error and square uniqueness)	276	3,649,041	13,221		

*Significant at the one per cent level of confidence.

level of confidence or better. None of the differences between the other conditions (*A*, *C*, and *D*) reached significance at the 5 per cent level for either of the two groups.

The analysis of the differences in performance with the combinations of planes of rotation of Group I and Group II is essentially a "between groups"/"within groups" analysis (3). The mean square for individual subjects was divided by the pooled error mean square, obtained by adding together the error sums of squares for the two groups, and dividing this by 276 (2 × 138). The resulting *F* was significant at the one per cent level. A further *F* was computed, the mean square for groups over the pooled mean square for subjects within groups. This ratio was significant at the one per cent level, indicating that the difference in performance between the two groups was significant, Group II having consistently higher scores than Group I.

Table 3 presents the means for each of the groups for the combinations of directions, the sequences of directions, and the ordinal position of the directions.

TABLE 3
MEANS FOR DIRECTIONS, SEQUENCES, AND ORDINAL POSITIONS IN SEQUENCE
(IN MILLIMINUTES)

Directions	*A*	*B*	*C*	*D*
Group I	367.3	467.7	390.0	384.5
Group II	464.4	542.0	455.1	437.2
Sequences	1	2	3	4
Group I	370.0	456.4	357.9	425.3
Group II	494.7	438.9	453.4	511.8
Ordinal position	1st	2nd	3rd	4th
Group I	337.7	373.7	481.7	416.3
Group II	373.5	478.5	557.6	489.2

TABLE 4
DIFFERENCES BETWEEN GROUPS I AND II FOR EACH OF THE COMBINATIONS OF DIRECTION OF ROTATION

Combinations of directions of rotation	Difference between Groups I and II	Significance level of the difference
A	97.04	1%
B	74.27	1%
C	65.08	1%
D	52.75	<5%>1%

The difference between the two Groups (I and II) for each of the combinations of direction of rotation was tested by means of the *t* test. Table 4, which summarizes these results, shows that the difference for each of the combinations was significant at the 5 per cent level of confidence or better, with Group II being superior in every case.

In order to estimate the relative efficacy of the combination of directions *B* for the two combinations of planes of rotation, the difference between *B* and each of the other three combinations (*A, C,* and *D*) for Group I was tested against this same difference for Group II. The resulting *t* did not reach significance at the 5 per cent level. This means that there is no evidence that the superiority of combination *B* over the other combinations was greater for Group I than for Group II, or vice versa.

It will be recalled that for two sequences of the latin square design there were two *complete* reversals of direction of rotation, while for two others there were no *complete* reversals. Therefore, it seems justifiable to average the means for each pair of sequences and test the difference between these two averages. This was done for both groups, and the resulting *t*'s showed that for Group I the difference was not significant at the 5 per cent level while for Group II the mean of the complete reversal sequences was superior at the one per cent level of confidence. Thus, on the basis of the available data, we can only conclude that the evidence for differential amounts of interference with complete reversal as compared to incomplete reversal is ambiguous.

G. Discussion

On the basis of the results reported above it is apparent that when the control arrangement is such that clockwise rotation of the left control moves the display to the right and clockwise rotation of the right control moves the display away from the subject, significantly superior performance is obtained to that with any of the other conditions tested. This superior combination of directions (Combination *B*) involves a continuity between planes of movement of the controls and direction of movement of the display. Another way of stating the results is to point out that under these conditions movement of each control produces the "expected" movement of the display. These results lend further support to an hypothesis proposed in the preceding study (7), namely that optimal operating conditions involve continuity between the plane of movement of a control and the resulting movement of a display.

The fact that there were no significant differences in performance among the other three combinations of directions used, although for two of them (*C* and *D*) the control movement for *one* hand produced the "expected" movement of the display while for the third (*A*) *neither* hand produced the "expected" movement, is of interest. It seems to indicate that as long as the optimum relationship is not maintained for *both* hands it is not important whether one or both the hands are performing under non-optimal conditions. This is probably a function of the fact that the scoring of performance on the SAM Two-Hand Coördination Test is such that it is necessary for *both* hands to perform "correctly," in order for the time-on-target score to accumulate. Thus, if the target follower is in line with the target in one dimension but not in the other, the score is the same as when the target follower is not in line with the target in either dimension.

Although previously (7) no difference was found between the two combinations of planes of rotation of Groups I and II, the present study provides an answer to the question of whether, when interference is a factor, the combination used in Group II will produce better performance. Not only did the overall performance of Group II exceed that of Group I, but for each of the combinations of rotation Group II was superior to Group I at the 5 per cent level of confidence or better.

In comparing Groups I and II with respect to the relative superiority of Combination *B* over the other combinations used (*A, C,* and *D*) a statistically insignificant *t* ratio was found. This means that the relative superiority of Combination *B* over the other combinations was not different for the two planes of rotation combinations used (Groups I and II).

From the results of this study and the previous study (7) further confirmation is gained for the suggestion that in the design of equipment it is important that the relations between movement of equipment controls and the display movement produced by these controls be such that there is continuity as well as the "expected" movement relationship between the two factors. In these two studies it has been shown that these relations hold when continuous, complex, two-hand coördination is the task, as well as in the simple tasks previously investigated. In other words, for the most efficient equipment design, the response of the operator, whether simple or involving complex coördination of the two hands simultaneously, should result in the "natural" or "expected" movement of the display or indicator, and further there should be a continuity in the planes of the control movements and the resulting display movements.

H. SUMMARY

The present study was designed to investigate performance of a two-hand coördination task as a function of the direction of rotation of the controls in relation to the direction of movement of the display. The SAM Two-Hand Coördination Test was modified so that (*a*) the effect of the direction of rotation of the crank on the movement of the display could be reversed and (*b*) the cranks could be placed in horizontal or vertical planes.

Two groups of 48 subjects each were given four trials, one on each of the experimental conditions described below. For Group I the two cranks were placed so that they rotated in a vertical plane parallel to the frontal body plane and for Group II the left crank was placed as for Group I, while the right crank was placed so that it rotated in a vertical plane perpendicular to the frontal body plane. The plane of rotation of the right crank was the only experimental difference between Groups I and II. The combinations of directions of rotation were as follows: (*a*) clockwise rotation of the left crank moved the target follower to the left and counterclockwise rotation moved it to the right; clockwise rotation of the right crank moved the target follower towards the subject and counterclockwise moved it away from the subject; (*b*) clockwise rotation of the left crank moved the target follower to the right and counterclockwise moved it to the left; clockwise rotation of the right crank moved the target follower away from the subject and counterclockwise moved it toward the subject; (*c*) the left crank as in (*b*) and the right as in (*a*); (*d*) the left crank as in (*a*) and the right as in (*b*). A latin square design was used in the presentation of the experimental conditions.

The results were analyzed in terms of the mean time on target for each

of the combinations of direction of rotation. Experimental Condition *B* (see above) was found to be superior to *A, C,* and *D* both for Group I and Group II. In neither Group I nor II was any of the differences between Conditions *A, C,* and *D* statistically significant. Group II was superior to Group I in terms of the overall estimate and in terms of each of the experimental conditions.

These results indicate that in the design of equipment controls superior performance will result if there is a continuity as well as an "expected" or "natural" movement relationship between the plane and direction of movement of the controls, and the plane and direction of movement of the display.

REFERENCES

1. CARTER, L. F., & MURRAY, N. L. A study of the most effective relationships between selected control and indicator movements. (Ch. 10 in Fitts, P. M. (*Ed.*), *Psychological Research on Equipment Design.*) Report No. 19, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.
2. GRANT, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.,* 1948, **45**, 427-442.
3. LINDQUIST, E. F. Statistical Analysis in Educational Research. Boston: Houghton Mifflin, 1940.
4. MELTON, A. W. (*Ed.*). Apparatus Tests. Report No. 4, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.
5. MITCHELL, M. J. H. Direction of movement of machine controls: III. A two-handed task in a discontinuous operation. Ministry of Supply, Psychological Laboratory, Cambridge, Nov. 1946 (FPRC No. 637 cont.).
6. ———. Direction of movement of machine controls: V. A two-handed performance in a continuous task. Applied Psychology Panel, Psychological Research Unit, Psychological Laboratory, Cambridge, May, 1949.
7. NORRIS, E. B., & SPRAGG, S. D. S. Performance on a following tracking task (modified SAM Two-Hand Coördination Test) as a function of the planes of operation of the controls. *J. of Psychol.,* 1953, **35**, 107-117.
8. REED, J. D. Factors influencing rotary performance. *J. of Psychol.,* 1949, **28**, 65-92.
9. VANRIPER, C., & BRYNGELSON, B. Speed and accuracy of clockwise and counterclockwise movements. *J. of Psychol.,* 1935, **1**, 247-253.
10. VINCE, M. A. Direction of movement of machine controls: I. Flying Personnel Research Committee, No. 637, Ministry of Supply, Aug., 1944.
11. VINCE, M. A., & MITCHELL, M. J. H. Direction of movement of machine controls: II. Ministry of Supply, Psychological Laboratory, Cambridge (FPRC No. 637 cont.), Nov. 1946.
12. WARRICK, M. J. Direction of movement in the use of control knobs to position visual indicators. (Ch. 9 in Fitts, P. M. (*Ed.*), *Psychological Research on Equipment Design.*) Report No. 19, U. S. Army Air Forces Aviation Psychology Reports. Washington, D. C.: U. S. Government Printing Office, 1947.

Department of Psychology
University of Rochester
Rochester 3, New York

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 131-142.

THE VALIDITY OF CERTAIN MEASURES OF ART APPRECIATION IN RELATION TO A DRAWING TASK*

Department of Psychology, Miami University

C. W. CRANNELL

A. INTRODUCTION

For a number of years there have been available to psychologists several instruments for the evaluation of factors in artistic performance. Well known are the Meier Art Judgment Test (5) and the Graves Design Judgment Test (3). It should be of some concern to psychologists that these two tests are administered to a great number of individuals every year at the secondary school and college level, and yield scores that, as the present writer can testify from personal observations, are employed (hopefully, with discretion) by counselors and teachers in giving suggestions and interpretations to the examinees. Any such suggestions would have to be based almost wholly on the face validity of the instruments. The norms provided by their authors are in actuality nothing more than a statement of such validity, for these norms are based on differentiation of samples already in art training from samples which have been subjected to little or no such training. It is therefore possible that these tests make discriminations based on the degree of sophistication in relation to works of art derived by taking courses from art teachers. A conclusion of this nature is reached by Barrett (1) in a study of the Meier Art Judgment Test and certain other instruments.

When a student is counseled against a career in design or other art work on the basis of poor performance on either of these instruments, it may well be that his low scores indicate a lack of experience, and that absorption of the proper experiences in art training would result in superior scores. The kind of research which would lead to a determination as to whether these tests have any predictive validity, or are merely achievement measures, must investigate their relationship to some measures of actual productivity. Meier's manual, and the prodigious amount of related research which he directed during the period 1926 to 1942, merely offer promising leads in this direction—the reader is referred to the bibliography in Meier (5, pp. 23-24).

---

*Received in the Editorial Office on August 31, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

It is possible that the highly controversial subjectivity and complexity of artistic performance deter many psychologists from attempting to reach precision in measurement of such work, at least at a level beyond elementary school. This problem is given a thorough treatment by Faulkner (2). Nevertheless there should be some of us to keep in mind the diligence of Fechner, the founder of experimental esthetics, and to keep alive an interest in the measurement of these complex human activities.

The present study is designed to indicate some preliminary avenues of approach to the task of providing more tangible evidence on the predictive efficiency of the two tests in question. The experiments to be reported began as a laboratory exercise in a course in experimental psychology. In 1947 the writer administered the Meier Art Judgment Test to a class of 41 psychology majors. It was found difficult to give the individual students any satisfactory statements about the validity of the test, particularly validity in terms of their own experiences. The writer therefore had each student draw, with a 20-minute time limit, a picture of a large, silver-plated elephant (Figure 1), and in a subsequent laboratory period had the students evaluate these drawings by the method of comparison by pairs. The correlation between the Meier Art Judgment Test scores and the standard scores derived from the evaluation was .327.[1] However, in this preliminary group the students all knew their own scores on the art test before they made their drawings. In the ensuing years, in which the data to be reported were collected, the drawings were made without such knowledge.

B. PROCEDURE

1. *Drawing of the Object*

At the opening of the first laboratory session in the second semester of experimental psychology (a required course for all psychology majors at Miami), the students, primarily juniors, who attended the laboratory in groups of six to eight at a time, were seated at large tables with cardboard barriers to prevent a student from seeing the work of another. The following instructions were read to them.

> We are going to see how well you can make a drawing of a specific object. Even though you think you cannot draw well, please try the very best you can. For the results of this experiment to be any good, it is necessary for the subjects to feel motivated to try to produce as fine a drawing as they possibly can. In a few minutes I shall put the

---

[1] A coefficient of .35 is reported by Barrett (1) for this test with combined ratings of four judges on several drawings made by pupils in Grade IX.

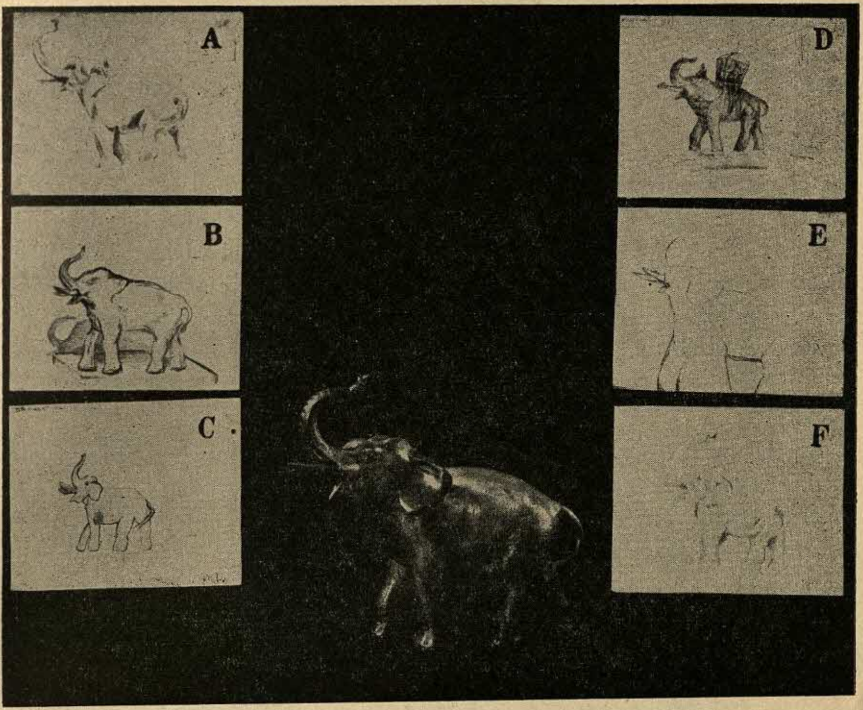


FIGURE 1

ELEPHANT USED AS A MODEL FOR THE DRAWINGS

Also shown are six of the drawings. *A, B,* and *C* are typical of items consistently given very high, average, and very low evaluations, respectively. Drawings *D, E,* and *F* are discussed in text. Drawing paper is 8½ x 11 inches.

object you are to draw up on this box. You will be given exactly 20 minutes in which to make your drawing. Make the drawing in pencil. Not only will you be allowed 20 minutes, but you are also supposed to use nearly all of this time in making your drawing. This instruction is necessary otherwise, some of you might give up in a short time because you feel that you cannot draw very well. You are supposed to take the entire 20 minutes, and to do your very best. You may draw the object just as it is or you may add appropriate details to the environment as well. In other words, you may employ as much artistic imagination as you think necessary to produce an esthetically good picture. On the paper which I shall now pass out do not put your name. Instead, place the paper with the long edge away from you like this and print "Experiment 32" in the upper left-hand corner. During the 20 minutes I shall provide you with a code number which is to be put in the upper right-hand corner. Are there any questions? . . . This is the object you are to draw.

At this point *E* took the elephant (Figure 1) from a place of concealment and placed it on a box which was draped in black cloth. The background was also black, and the object was lighted from about 4 ft. away and 45° above by a 200-watt lamp. Since the *S*s were seated at two rows of tables, the nearest *S* was about 8 ft., the farthest about 13 ft. from the object. Time elapsed was told the *S*s at 5, 10, 15, and 18 minutes. At exactly 20 minutes *E* said "Stop drawing now," and collected the drawings.

The Meier Art Judgment Test was then administered without time limit. Beginning in 1949, this test was followed by the Graves Design Judgment Test, administered without time limit. The tests were then scored and checked by students after exchanging answer sheets.

2. *Evaluation*

At the next laboratory session, a week later, the drawings were evaluated. Since these procedures varied from year to year, it is necessary to describe each group of students separately.

Group 1 (1948; 49 students). At opposite sides of the room two placards were mounted. Each placard contained 10 drawings made in 1947. Each set of 10 was scaled on the basis of the comparison by pairs treatment made in 1947. The proportions of preference had been transformed into standard scores, following an adaptation of the method suggested by Guilford (4, p. 236). The selected items were scaled approximately .20 *SD* apart. Each *S* prepared an answer sheet containing the code numbers of the drawing on each side. Half of the *S*s sat before each placard. The following instructions were read:

> We are going to rate the drawings you made last time by comparing them with a series of drawings made last year, and scaled in approximately equal steps from very poor to very good. Each *S* will receive a stack of approximately equal size of the drawings. Take the drawing from the stack, turn it over, and compare it with the placard. Decide which of the drawings it most nearly resembles as an artistic production. Then find the code number of the drawing on your answer sheet and place the number of the drawing on the placard after it. If you think the drawing in your hand is less artistic than No. 1 on the placard, place a zero after its code number; if you think it is more artistic than No. 10, place an 11 on your answer sheet after the code number.

Further instructions were then given on passing the drawings along to the next student. After all the drawings had been rated, the students exchanged places before the placards and again rated all drawings. All drawings were shuffled between laboratory sections. From these data it was possible to obtain a mean rating on the 12-point scale for each drawing, based on 48 judgments when the self-rating of each student's own drawing was excluded, for each placard.

Group 2 (1949; 34 students). Each *S* was instructed to prepare an answer sheet on graph paper. This answer sheet was ruled off into a plot with the code numbers of 14 of the drawings along the left side and along the top. The following instructions were given:

> Today we shall rate a number of the drawings using the method of comparison by pairs. In this section we shall judge 14 drawings, each of which has one of the code numbers shown on your answer sheet. You are to compare each drawing with every other drawing and indicate in each case which of the two you like better. I shall now give this stack of drawings to the first *S*. He (she) will take the first one and use it as a model for comparison with all the remaining 13. The code number of this drawing will determine the column in which the first 13 entries are placed. He (she) will compare this drawing with the next in the stack and decide whether he (she) prefers the model, that is, the first drawing, or this second one. If the model is preferred, put the letter *"P"* in the proper space; if the second drawing is preferred to the model, put the letter *"N"* in the space. He (she) now passes this second drawing on to the second *S*, who retains this drawing for a model, and waits for the third drawing to make his (her) first comparison. In other words, you keep the first drawing you receive as a model and pass the rest along.

Further instructions clarified the procedure for picking up a new model when all 13 comparisons had been made, until each of the 14 drawings had served as a model for each *S*.

The pictures used in each laboratory section were those drawn in other sections, so that the *S*s had never seen any of the pictures before. Since there were six sections, each picture was evaluated at least twice, and a different combination of 14 pictures was possible for each section. Two separate proportions of preference could thus be computed for each picture. The proportions were then transformed into standard scores.

Group 3 (1950; 37 students). In this group the general procedure of Group 1 was repeated. However, because one of the placards of 10 drawings used in Group 1 had not been wholly satisfactory, a new placard of 10 drawings, selected from the data of Group 1, was made up as a replacement. Furthermore, this new group rated both their own drawings and the 34 drawings made by Group 2, all 71 drawings being mixed at random.

Group 4 (1951; 31 students). In this group the general procedure of Group 2 was repeated, except that 16 instead of 14 drawings were compared in each section. Furthermore, 9 of the drawings selected from the data of Group 3 were included as "anchor" material.

Group 5P (1952; 23 psychology majors) and Group 5A (1952; 31 art majors). In these groups the general procedure of Group 2 was repeated. Group 5A was made up of 31 art majors, enrolled in the second semester of a required two-semester psychology course. Group 5P judged 20 pictures, 9 drawn in different sections of the same group, 11 drawn by Group 5A. Group 5A, which could meet only for an hour at a time, judged 14 pictures, 8 drawn in different sections of the same group, 6 drawn in 5P. The proportions of preference, for purposes of the present study were computed for the drawings by psychology majors only, and converted into standard scores, separately for Groups 5A and 5P.

In 1952 an additional evaluation of 194 of the drawings was performed independently by three experts, using a variation of the method of equal appearing intervals.[2] The expert was provided with a large surface for sorting the pictures, and given the following instructions.

> Here is a stack of 194 drawings of this silver elephant, most of them made by psychology majors, but a few by art majors. These drawings were made under a 20-minute time limit after the following instructions had been read (here the instructions to students for drawing the picture were read). You are to sort them into 9 piles from best to worst. The pile at this end of the tables should contain the best drawings; the pile at that end, the worst drawings. Pile 5, in the middle should contain drawings halfway between the best and the worst. The remaining piles, 8, 7, 6, and 4, 3, 2, should contain drawings which are in equal steps of quality along the scale, like steps of a stairway. The best way to begin is to spread out a dozen or so pictures from the stack and get an idea of the range of quality before you begin to build the nine piles. You may make as many changes from pile to pile as you wish to.

After sorting once began, no statements were made which might influence the nature of judgments. If the expert asked whether or not he might employ some particular variable, such as "originality," he was answered, "Use any criteria you wish in making your judgments." All three experts showed very conscientious behavior, and made many revisions as they worked. Each one, without being asked to do so, made a final scrutiny of the items in each pile before stopping.

---

[2] Of the 205 drawings collected in these groups, 10 had been used in constructing a scaled placard and one was a second drawing by a student repeating the course. The writer is deeply indebted to the three artists who each devoted well over an hour to the arduous task of judging 194 drawings: Miss Betty Woodbury, Mr. Allan H. Clarke and Mr. Edwin L. Fulwider.

C. RESULTS

1. *Evaluation by Experts*

For reasons of clarity of exposition, the most recently acquired data will be reported first. Table 1 presents the data on sum of judgments by the three experts on the nine-point scale of equal appearing intervals. This distribution is seen to have a severe positive skew. This same tendency is very strong in the judgments of each expert, taken alone. Each expert, while sorting the pictures, made the same complaint, that most of them were very poor indeed, and that further differentiation among these inferior pictures was not possible. Because of this positive skew, all correlations to be

TABLE 1
SUM OF JUDGMENTS BY THREE EXPERTS OF DRAWINGS OF AN ELEPHANT

Sum of three judgments	All drawings	Drawings by art majors	Drawings by psychology majors		
			All	Those with art training	No art beyond grade school
27	2	2			
26	1	1			
25	1		1	1	
24	3	3			
23	2	1	1	1	
22	2	1	1	1	
21	4	4			
20	3	2	1		1
19	3	1	2	1	1
18	2		2	1	1
17	3	2	1		1
16	7	2	5	3	2
15	7	1	6	3	3
14	7	1	6	3	3
13	8	3	5	2	3
12	1	1			
11	13	4	9	1	8
10	5		5	2	3
9	8		8	3	5
8	19	2	17	3	14
7	22		22	4	18
6	23		23	1	22
5	13		13	1	12
4	21		21	7	14
3	14		14	2	12
Sum	194	31	163	40	123
Mean	9.72	17.61	8.22	10.56	7.46
Crude Mdn	8	17	7	9.5	7

reported for the expert judgments are based on somewhat triangular scattergrams, which are, however, linear.

The third column of the table shows the subgroup of art majors, whose drawings were clearly discriminated, in the main, from those of psychology majors. The group of psychology majors is further broken down in the last two columns to show the differentiation of those who stated they had had art courses either in high school or college, or both, from those who claimed no training beyond grammar school, or none at all. Art training beyond grammar school is an elective subject, and may tend to be taken more often by those students who have some competence. The relevant data in the table suggest that either this idea is correct, or that the kind of art training given these students led to a somewhat better performance in drawing the elephant.

In the present study all further treatments will be confined to the data on psychology majors. The results for art majors will be the subject of a further report when more cases have been collected.

An estimate of the consistency among the three experts may be obtained by intercorrelating the ratings. The correlations between judgments of Experts 1 and 2, Experts 1 and 3, and Experts 2 and 3 are .550, .579 and .438, respectively. Applying the correction for broad categories from the table in Peters and Van Voorhis (6, p. 398), the coefficients become .576, .601 and .454. The average of these three coefficients, obtained via transformation into Fisher's z, is .547. If it is assumed that each rater contributes equally to the sum of ratings, this average r may be entered in the formula for correcting a reliability coefficient by tripling the length of the test. This yields .784, which may be taken as an approximation of the reliability of the sum of three ratings.[3]

For a study of validity of the tests in relationship to these evaluations, the data of Groups 2, 3, 4, and 5P only will be considered, since it was with Group 2 that the Graves Design Judgment Test was introduced. In these groups there are 123 cases who took both tests. The correlation between the sum of expert judgments and the Art Judgment Test is .369, between expert judgments and the Design Judgment Test, .172, and between the two tests, .438. These coefficients yield a multiple R between expert judgment and the two tests of .369, or .345 after application of

---

[3] A number of assumptions in this paragraph may be subject to criticism, especially the type of correction for broad categories applied in this case to skewed data. If the raw coefficients are taken, the average r is .525, which raises to .768 by tripling the test. Either treatment must be recognized as affording rough estimates.

Ezekiel's correction (Guilford, 4, p. 384). Beta for Art Judgment is .363, for Design Judgment .013.

Some doubt may arise as to the representative nature of this sample of psychology majors. Their mean and *SD* for Art Judgment are 98.32 and 8.87, respectively. Meier's manual (5, p. 19) shows a "raw score" mean of 83.74 for college, non-art students. This is quite evidently based on the actual number of items correct, rather than upon scores obtained with the key, which weights 25 of the 100 items; a comparison of the table containing the cited mean with the percentile rank norms on pages 10 and 11 of his manual makes this clear. These norms run about 18 points higher than the means in his table would indicate. Adding 18 to the mean above makes it quite similar to that of the present study. The mean and *SD* for Design Judgment are 57.97 and 15.32 respectively. This mean value is somewhat higher than the 49.40 given in Table 1 of Graves' manual (3), but the three samples used in his norms are all freshmen in engineering, dentistry, and business. One of these groups, 100 freshmen retailing students, has the mean of 57.75.

It was mentioned above that there was some relationship between the evaluation of the drawings and previous art training in high school or beyond. In the group of 123 students now under consideration, 30 stated that they had had such training. Biserial correlations of this differentiation with expert judgment, Art Judgment, and Design Judgment are .366, .393, and .445, respectively. A multiple *R* of expert judgment with the three variables, Art Judgment, Design Judgment, and stated art training is .443, or .420 after application of Ezekiel's correction. The Betas for the three variables in the order stated are .280, —.082, and .292. Partial correlations, holding art training constant, between expert judgment and the two tests are .263, slightly below the 1 per cent level of confidence, and .028, not significant.

The data thus far reported suggest that whatever factors are found in these drawings related to the appreciation of design and also influential in affecting their evaluation by experts are probably learned in formal courses in art. On the other hand, there seem to be factors measured by the Art Judgment Test which are related to expert evaluation of artistic production and at the same time independent of past training.

2. *Evaluation by Students*

One pervading feature of these data is the high degree of consistency found within student ratings and between student and expert evaluations, regardless of the system of evaluation employed. For Group 3, in which the data of both Groups 2 and 3 were rated, the correlation between ratings on one panel and ratings on the other is .958, corrected for split halves. The correlation between expert judgment and average rating on both panels is .717. For Group 4, the correlation between one set of comparisons and the other set, by entirely separate judges and with different combinations of the drawings, is .978, corrected for split halves. The correlation between the comparisons made in Group 2 (standard score form), and the ratings against the panels made in Group 3, for the drawings of Group 2, is .902, uncorrected. The ratings of the 9 drawings from Group 3 used as "anchors" in Group 4 correlate .967 (rho) with the comparisons involving these drawings made by Group 4. In Group 5P, the correlation of comparisons made by psychology majors with those by art majors is .859. In this same group, the correlation of expert judgments with psychology majors' comparisons is .680, with art majors' comparisons it is .520. If all standard scores from Groups 2, 4, and 5P are pooled, making a subsample of 96 cases, the correlation between these scores and expert judgment is .675.

With regard to validity of the two tests, the ratings made by Group 3 correlate .53 and .43 with Art Judgment and Design Judgment, respectively. For the comparisons made in Group 4, these coefficients are .41 and .15. The most complete analysis is possible using the standard scores in the subsample of 96 cases from Groups 2, 4, and 5P. Here the two correlations with Art and Design Judgment are .587 and .452, respectively. The correlation between the two tests is .494, and these data yield a multiple R with student judgment of .616, or .606 after application of Ezekiel's correction. The biserial correlations of previous art training in high school or beyond with each of these three main variables are .333, .374, and .580. The corrected multiple R in this case is .596. This indicates that there is no gain in prediction of student comparisons of the drawings through the inclusion of stated previous training.

A further possibility was explored: that of combining the expert and student judgments. To this end the sum of ratings by three experts was transformed into standard score form and added to the standard score from student comparisons for each case in the subsample of 96 cases. The two correlations of this measure with Art and Design Judgment are .533 and .421, slightly lower than those for student judgment alone. It is possible that some greater refinement of the method of combining the two kinds of evaluation would lead to a higher relationship, but this has not been undertaken. One of the reasons for a lower relationship between the tests and expert evaluation, as compared with these relationships to student evaluation, is without doubt the lower reliability of the expert judgments. Use of more than three experts would probably serve to strengthen these relationships. It must also be realized that the three experts averaged about an hour and a half in judging 194 pictures, which comes to about 1.4 minutes per picture. On the other hand, in a comparison by pairs experiment a single student, in comparing one drawing with 16 to 20 others, will spend perhaps 5 minutes per picture; since each drawing is seen by at least 12 students, this means that around 60 minutes are devoted to the evaluation of each item.

D. Summary and Conclusions

Over the course of six years college students in small groups were instructed to draw, with a 20-minute time limit, a picture of a large silver-plated elephant. Conditions of drawing were carefully standardized. A total of 194 such pictures was available for analysis. To 123 of these students both the Meier Art Judgment and the Graves Design Judgment Tests were administered. Various subsamples of these students performed evaluations of different sets of these drawings, employing in some cases the method of comparison by pairs and in others a rating against panels of 10 of the drawings which had been selected from the data in other subsamples. In addition, three competent artists, termed the experts in this report, evaluated all 194 drawings according to the method of equal appearing intervals. A high degree of consistency in evaluating these drawings was observed for various treatments, leading to the conclusion that the reliability of student judgments is somewhat over .90, the reliability for the three experts combined a little under .80.

The correlation between student judgment and scores on the Meier Art Judgment Test was found to be .587 for the largest subsample available (96 cases); for the same group, student judgment correlated .452 with the Graves Design Judgment Test. Expert judgment was found to correlate .369 with Art Judgment and .172 with Design Judgment. An additional finding was that any relationship between such expert evaluation and Design Judgment seemed to be contributed by past training of the students in art.

Of general significance is the fact that substantial relationships may be discovered between standardized tests of esthetic appreciation and a 20-minute sample of performance in drawing. Without doubt, the magnitude of such relationships depends heavily on the care taken and time expended in evaluating the drawings.

The present study leaves many questions unanswered. Was it best to use a three-dimensional model, or would a photograph have done as well? Would more, or less, complex drawing tasks yield better results? What gain would there be from strengthening the criterion by the addition of more drawing tasks, and, incidental to this, what is the reliability of performance measured by evaluation of a later repetition of the same task? The reliabilities reported in this study are actually based on consistency of judgment, not on consistency of performance. What about the item validities of the art appreciation tests? Finally, would such a measure of performance, included in a test battery, contribute to prediction of success in advanced training in art? The time and facilities of the present writer clearly limit his excursions into this area, and he would welcome the extension of this work by colleagues who, like Fechner, yearn to provide precision in those areas of human behavior where conjecture rules.

REFERENCES

1. BARRETT, H. O. An examination of certain standardized art tests to determine their relation to classroom achievement and to intelligence. *J. Educ. Res.*, 1949, **42**, 398-400.
2. FAULKNER, R. Evaluation in art. *J. Educ. Res.*, 1942, **35**, 544-554.
3. GRAVES, M. Design Judgment Test, 1948 Manual. New York: Psychological Corporation, 1948.
4. GUILFORD, J. P. Psychometric Methods. New York: McGraw-Hill, 1936. Pp. 566.
5. MEIER, N. C. The Meier Art Tests: I. Art Judgment, Examiner's Manual. Iowa City: Bureau of Educational Research, University of Iowa, 1942.
6. PETERS, C. C., VAN VOORHIS, W. R. Statistical Procedures and Their Mathematical Bases. New York: McGraw-Hill, 140. Pp. 516.

Department of Psychology
Miami University
Oxford, Ohio

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 143-153.

A METHODOLOGICAL STUDY OF SIZE-DISTANCE PERCEPTION*

Department of Psychology, Princeton University

WILLIAM M. SMITH

A. INTRODUCTION

While few, if any, psychologists would question the general principle of perceptual constancy in behavior, the scrutiny of more specific "constancy" phenomena would not meet with such general acceptance. This statement is particularly true with regard to the problem of "size-constancy" in visual perception. This problem has been investigated widely, yet there is much disagreement about it with respect to certain aspects of methodology and experimental findings. For example, different methods have been used in various studies (4, 5, 7, 8, 9, 10, 12) to determine how perceived size varies with physical distance. Results from these methods often have been disparate. It has been pointed out by Joynson (9), for example, that the method used by Thouless (12) precluded getting results which could ever demonstrate perfect "constancy." It is thus easy to see how such a methodology might lead to that view of size perception which is conceived to be a kind of compromise process. This particular point of view has its more recent complement in the theory of size proposed by Gilinsky (7) which conceives the perception of size as falling between two limiting "laws," the "law of size-constancy" on the one hand, and the "law of the visual angle" on the other. While Gilinsky maintains that relevant data from size-distance experiments support her formulation, it has been pointed out by Smith (11) that all the evidence is not compatible with this position. Such incompatible evidence comes from studies (4, 5, 8) which have shown, at least for binocular vision, that perceived size in some cases *increases* with distance rather than decreases or remains invariant. These data, if not artifacts, do violence to the usual conception of "size-constancy" and must be integrated with other data on size-distance relations. One interest of the present study, utilizing three-dimensional objects, was to determine whether a "natural" rather than a laboratory situation would give rise to this kind of evidence.

---

*Received in the Editorial Office on August 31, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

Another problem about which there is little agreement among investigators in this area deals with the measurement of apparent or perceived size. The methods which have been used are nearly as abundant as the various definitions of "size." In general, the most common method employed is one in which the subject is presented with a standard object at some distance and a comparison object usually near at hand. The task for the subject is to choose or alter the comparison object until the standard and comparison are judged equal in size. There seems to be no serious objection to this method except that the experimenter's definition of equality may be equivocal to the subject. In one case equality may be taken by the subject to mean that he is to choose a comparison which *looks* the same size as the standard; in another case, it may be taken by the subject to mean that a comparison be chosen which is the same physical size as the standard. The problem involved here is probably the "paradox of converging parallels" referred to by Boring (1). Another purpose of the present study was to investigate this common procedure of measuring perceived size by employing two different conditions of observation. One condition followed the typical procedure; one of a series of comparison objects nearby was matched with a standard object at some remote distance. In the other condition this arrangement was reversed; one of a series of comparison objects at a remote distance was matched with a standard object nearby. It was felt that any systematic difference found between these two conditions would require a more serious analysis of the common procedure for obtaining size judgments in size-distance studies.

Distances of greater magnitude than those found in other studies (with the exception of Gibson, 5) were used in the present study. Data for considerable distances are badly needed. One of the purposes of this investigation was to obtain such needed data.

B. SUBJECTS

The subjects were six college freshmen with 20/20 vision (3 corrected) and naïve regarding the purpose of the experiment. No subject was acquainted with formal knowledge about size-constancy behavior.

C. STIMULUS MATERIAL

The stimulus material consisted of wooden cubes painted white. Each cube was mounted on a brass rod which was inserted in the ground for purposes of presentation to the subject. Standard blocks, two inches and four inches in size, were used. A series of comparison blocks were made up which varied in size from one-half inch to six inches in one-quarter inch steps.

A table-like structure with a small black curtain was constructed. The subject sat behind this structure when making observations. The surface of the table was about chest-high to the subject when seated. The standard and variable objects were presented on the front end of the table at a distance of approximately two feet from the subject's eyes. Between each trial the curtain was drawn so that the subject could not observe the changing of the stimulus objects.

D. PROCEDURE AND EXPERIMENTAL DESIGN

Observations were obtained under two conditions. (*a*) *Condition N*: The standard stimuli were placed at some remote distance and the comparison stimuli nearby. (*b*) *Condition R*: The comparison stimuli were placed at some remote distance and the standard stimuli nearby. The near distance was always that distance from the subjects' eyes to the front edge of the table (approximately two feet). The far distances used were 16 ft., 80 ft., and 320 ft. Every subject observed under both conditions. Three subjects were given Condition *N* first, the other three Condition *R* first. Two experimental sessions were required of every subject, one for each of the two conditions.

A constant procedure and order of presentation was used for all subjects and both conditions. The three viewing distances, the two standard stimuli, and an ascending and descending series of comparison stimuli were arranged in a random order of presentation. Three different orders of presentation were used, each consisting of 12 trials each (3 x 2 x 2). By this procedure a mean of three judgments for each standard object was obtained under each of the three distances and the ascending and descending order of presentation. For the three trials involving the same standard and the same viewing distance the starting point in the comparison series was different on each trial. These starting points were determined at random.

The procedure for Condition *N* was as follows. The subject was seated behind the table-like viewing apparatus with the curtain drawn. The following instructions were read to him:

> I am going to show you a white block which will be at some distance directly in front of you. Then later I will show you another white block a little to your right (pointing to the front edge of the table). Your job is to look at the blocks and decide whether the one on the edge of the table is bigger or smaller in physical size compared

> to the block in the distance. If it is smaller, then you simply say "too small," if larger simply say "too large." Whenever the two blocks look alike in physical size, say "equal." By saying they are equal in physical size you should mean that if both blocks were measured by a ruler they would have the same dimensions in inches. If you should ever say "too small" for example, and then find the next block "too big," you may choose between these two blocks, that is, between the block seen last and the one just before it. In all cases you must make a choice. Do you have any questions?

The instructions for Condition *R* were the same as described above, except that it was explained that a block would be shown at the front edge of the table and the task was to judge whether the block seen in the distance was too big, too small, or equal in physical size compared to the block at the edge of the table.

The experiment required two experimenters.[1] When testing under Condition *N*, one experimenter stood at the right of the table and presented the variable stimuli (or standard for Condition *R*) one by one by holding the rod attached to the block. The subject could not see the rod. The second experimenter changed the size and distance of the remote stimuli as required. He was never in view when the subject made judgments. Between presentations of the comparison stimuli for Condition *R* the subject would drop his head and close his eyes while the experimenter removed the comparison stimulus and put the next of the series in its place. After a judgment of equality was made the curtain was pulled and the subject waited for the next trial. The subject was given a brief rest after each series of 12 trials.

The stimulus objects were placed so that they rested about one inch from the surface of the lawn. Each block was oriented in the same manner, i.e., the surface facing the subject was rotated approximately 45° to the subject's left. The same orientation was given to the blocks presented at the front end of the table.

The experiment was performed outdoors on a large flat lawn bordered on one side by a road and trees. The background consisted of additional bushes and trees.

E. Results

Tables 1 and 2 give the results of the experiment for every subject on both standard objects under the two conditions of observation. Each entry in the table is the mean of six trials, three for the ascending and three for the descending order of presentation of the comparison objects. Although

---

[1] The author is indebted to Mr. R. B. Altmayer for his assistance in this experiment.

TABLE 1

SUMMARY OF SIZE JUDGMENTS FOR CONDITION *N* WITH AVERAGE DEVIATIONS (ALL ENTRIES IN INCHES)

	16 Ft.				80 Ft.				320 Ft.			
	2 in.		4 in.		2 in.		4 in.		2 in.		4 in.	
Subject	X	AD	X	AD	X	AD	X	AD	X	AD	X	AD
1	2.12	.12	3.96	.15	2.42	.22	4.08	.17	2.67	.33	4.58	.28
2	1.79	.14	3.54	.14	1.96	.30	4.04	.39	2.17	.33	4.08	.39
3	2.12	.12	3.92	.19	2.17	.17	4.25	.42	2.58	.44	4.42	.22
4	1.96	.07	4.08	.22	2.04	.21	4.21	.30	2.46	.46	4.67	.17
5	1.88	.12	3.62	.12	1.79	.07	3.67	.17	1.71	.14	4.38	.46
6	2.25	.17	4.17	.25	2.67	.19	4.67	.33	3.42	.31	5.25	.17
Grand mean	2.02		3.88		2.18		4.15		2.50		4.56	

TABLE 2

SUMMARY OF SIZE JUDGMENTS FOR CONDITION *R* WITH AVERAGE DEVIATIONS (ALL ENTRIES ARE IN INCHES)

	16 Ft.				80 Ft.				320 Ft.			
	2 in.		4 in.		2 in.		4 in.		2 in.		4 in.	
Subject	X	AD	X	AD	X	AD	X	AD	X	AD	X	AD
1	2.04	.07	4.17	.11	2.04	.15	4.17	.19	1.88	.17	3.96	.29
2	2.50	.25	4.58	.17	2.33	.25	4.50	.33	2.17	.58	4.58	.36
3	2.00	.08	3.67	.11	1.92	.11	4.12	.33	2.00	.25	3.75	.25
4	2.21	.29	4.21	.29	1.67	.28	3.92	.44	1.58	.42	3.58	.31
5	2.21	.21	4.38	.25	1.92	.28	4.29	.47	1.96	.28	3.92	.36
6	2.04	.21	4.04	.29	1.58	.33	3.83	.50	1.42	.33	3.17	.42
Grand mean	2.17		4.18		1.91		4.14		1.84		3.83	

not shown in the tables, a constant error due to order of presentation of the comparison objects was apparent in the data, i.e., measures of equality for the ascending order were in most cases less than similar measures for the descending order. Also given in the tables is the average deviation for each set of six trials. It can be seen from the tables that in nearly every case the average deviation increases as distance increases. This finding is in line with other studies of perceived size involving considerable distances (5, 8).

The effect of the distance variant is apparent from the data in the tables, but can be seen more clearly in the graphs of Figure 1. In the graphs the distance in feet of the standard object (or comparison object in Condition *R*) is plotted against the size in inches of the comparison object. Every point on each curve is the mean of six trials. Curves are drawn for every subject on both standard objects and the two conditions of observation. In addition, similar curves are presented for grouped data. These curves are

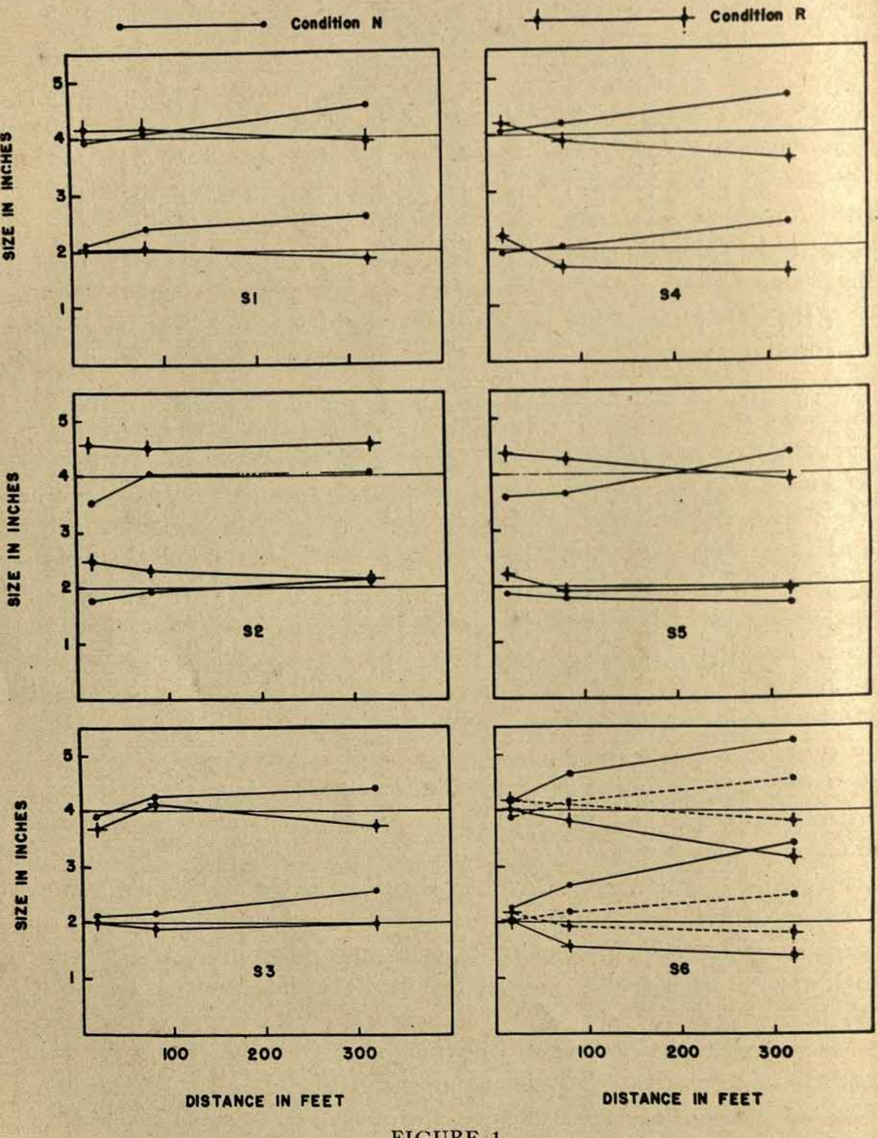


FIGURE 1
Perceived Size as a Function of Distance for Condition *N* (Normal) and Condition *R* (Reverse)

shown on the graph for Subject six. Inspection of the curves in Figure 1 shows the effect of the distance variant upon perceived size. In 11 out of 12 comparisons (for Condition *N*) an increase in the distance of the standard object is associated with an increase in size of the comparison object. This relation holds true for every subject except Subject 5 on the two-inch standard. Taking the physical size of the comparison object as the measure of perceived size, these data say, in effect, that perceived size *increases* with distance. As the physical distance of the standard object increased the subject required a larger and larger comparison object for a judgment of equality. The data for Condition *R* are, in general, consistent with the data of Condition *N*, although these curves show that as physical distance increases, the size of the comparison object chosen for equality decreases. However, it should be remembered that in Condition *R* the distance of the comparison objects was varied from trial to trial, and the standard object remained fixed at the near viewing distance. If perceived size increased as physical distance increases, as the data of Condition *N* show, then one would expect that a smaller and smaller comparison object would be required to match a given standard object *as the distance of the comparison object increased.* The curves of Figure 1 show by and large that this expectation is confirmed, although in three of 12 comparisons perceived size seems to be invariant with distance. A simpler way of comparing the two conditions is to visualize the symmetry of the pairs of curves. If the two conditions gave equivalent data, then one curve would be the mirror-image of the other. This symmetry is particularly apparent for the curves of the grouped data. It is of importance to note that in only one instance out of 24 did perceived size, as measured, *decrease* with physical distance.

In general, the pairs of curves also show, with the exception of Subject 2 and Subject 5, that there is substantial agreement between the size of the comparison object and the standard object at the nearest viewing distance (16 ft.). The data for this distance show reasonable "size constancy." While the data of Subject 2 and Subject 5 show the general effect of the distance variant, the agreement between the size of the comparison object and the standard object is poorest at the nearest distance. In terms of "size constancy," extrapolation from their curves would indicate that for them better "constancy" is obtained at considerably greater distances. It is possible that the factor involved here is that of "normal viewing distance," a factor which would be expected to vary with the observer.

F. Discussion

Although the results show that the agreement within and among subjects on the two conditions of observation is far from perfect, the similarity between conditions overall is good enough to indicate that it makes little difference whether the distance variant refers to the standard object or to the comparison object in this experiment. In both cases, the effect of distance was similar, that is, as distance increased the perceived size, as measured by these procedures, also tended to increase. For the two subjects who showed the least agreement between the two conditions, there was a general tendency to choose comparison objects too large in size for the nearer distances under Condition *R* and too small in size under Condition *N*. This tendency was not characteristic of other subjects, and probably reflects not a difference between the two conditions, but an underestimation of near distances under both conditions for these subjects.

One should, of course, expect the effect of the distance variant on perceived size to be the same for any object, if the effect is a real one. The conclusion fom the data of Condition *N* is that perceived size increases as distance increases. This in turn indicates that if a comparison object increases in distance from the subject its perceived size should also increase. Thus for a nearby standard object of a particular size, a smaller and smaller comparison object would be needed for a match as the distance of the comparison increased. In so far as subjects under Condition *R* tended to choose smaller and smaller comparison objects as their distance increased, the interpretation is justified that perceived size increased with distance and the arrangement of comparison and standard object made little difference in general.

The more provocative aspect of the present results is not the comparability of the two conditions but the fact that perceived size, as measured, increased with distance rather than remained constant or decreased. This finding is contrary to the usual experimental results on "size-constancy" and size-distance relations and alone might be dismissed as due to an experimental artifact. However, such results have been reported by Holway and Boring (8) (who did dismiss their results on the basis of a space error), by Gibson (5), and by Chalmers (4). In all these studies, binocular vision was the condition of observation under which such results were obtained. Furthermore, the three studies used distances considerably greater than had been used in previous systematic studies. Gibson (5) used distances up to one-half mile, although the acceleration of perceived size with distance was only apparent between approximately 80 feet and 675 feet, being greatest between 80 feet and 320 feet. Chalmers used much the same procedure as Holway and Boring, and while his data showed the same sort of increase in perceived size for binocular vision, the size-distance function was not linear as in the Holway and Boring data, but began to taper off at about 30 ft. It should be noted that these two studies, in particular the one by Chalmers, presumably offered no cues in the visual field, but only visual cues, such as disparity, accommodation, and convergence. While the space error hypothesis of Holway and Boring is not unreasonable, "over-constancy" in their study, in Gibson's study and in the present one, make it somewhat gratuitous to assume that a space error is operating. This is particularly true with regard to Gibson's results. For the present a more direct interpretation of the effect is preferred.

This interpretation involves the proper estimation of distance. It is a generally accepted fact that the correct perception of size depends upon an accurate perception of distance (2, 3, 6). If distance perception errs in the direction of over-estimation, then perceived size will be greater than it would be if distance perception were accurate. If the error is in the direction of underestimation, perceived size will be too small. From this relationship it can be inferred that subjects in the present experiment tended to systematically overestimate the distance of the standard objects (or comparison objects in Condition *R*). Furthermore, the overestimation must have been proportionally greater as the actual distances increased, in order to account for the acceleration of the size-distance functions. In the absence of independent measures of perceived distance, however, this interpretation is very tentative and would require such independent distance measures for confirmation.

It can thus be seen that rather widely different experimental situations have given results which are very similar in nature. The two previous studies which found an increase in perceived size with distance for binocular vision (4, 8) were straightforward laboratory studies, utilizing two dimensional stimuli and a restricted visual field. The other study (5) and the present one were carried out in a natural setting using three-dimensional objects. Furthermore, instructions were quite different. Until proven otherwise it is assumed that the factor common to all of these studies was the overestimation of distance. This interpretation, if acceptable, does no violence to the usual conception of "size-constancy" and it does accommodate data which would be considered anomalous otherwise. Future experiments of size-distance relations must deal with independent measures of perceived distance, an approach evident in the recent work of Gilinsky (7).

G. SUMMARY

In the present experiment on the size-distance function two conditions of observation were compared: (*a*) Condition *N* involved the matching of nearby comparison objects with more remote standard objects, and (*b*) Condition *R* involved the matching of nearby standard objects with more remote comparison objects. White cubes were used as stimulus objects. The standard objects were two inches and four inches in size. The comparison objects varied in size in ¼-inch intervals. The experiment was carried out in an outdoor environment. Observations under both conditions were obtained from six subjects. Three different distance values were investigated: 16 feet, 80 feet, and 320 feet. On the basis of the comparison between the two conditions of observation and the analysis of the size judgments with respect to distance the following conclusions were reached:

1. It makes little difference whether the standard object or the comparison object is varied in distance in such a study as the present one.
2. The data for both conditions showed that perceived size, as measured, increased with distance rather than decreased or remained invariant.
3. It was tentatively concluded that these results of the present study and similar results from other studies were due to errors of localization and not to an experimental artifact.
4. Future studies on size-distance relations should investigate independent criteria of perceived distance before a more secure interpretation can be given to the wealth of data on "size-constancy" and size-distance perception.

REFERENCES

1. BORING, E. G. Visual perception as invariance. *Psychol. Rev.*, 1952, **59**, 141-148.
2. ———. Sensation and Perception in the History of Experimental Psychology. New York: Appleton-Century, 1942.
3. CARR, H. An Introduction to Space Perception. New York: Longmans, 1935.
4. CHALMERS, E. L. Monocular and binocular cues in the perception of size and distance. *Amer. J. Psychol.*, 1952. (In press.)
5. GIBSON, J. J. Motion Picture Testing and Research. AAF Research Report No. 7. Washington, D. C.: U. S. Government Printing Office, 1947.
6. ———. The Perception of the Visual World. Boston: Houghton Mifflin, 1950.
7. GILINSKY, A. S. Perceived size and distance in visual space. *Psychol. Rev.*, 1951, **58**, 460-482.
8. HOLWAY, A. N., & BORING, E. G. Determinants of apparent visual size with distance variant. *Amer. J. Psychol.*, 1941, **54**, 21-37.

9. JOYNSON, R. B. The problem of size and distance. *Quart. J. Exper. Psychol.,* 1949, **1**, 119-135.

10. LICHTEN, W., & LURIE, S. A new technique for the study of perceived size. *Amer. J. Psychol.,* 1950, **63**, 280-282.

11. SMITH, W. M. An analysis of Gilinsky's theory of visual size and distance. *Psychol. Rev.,* 1952, **59**, 239-243.

12. THOULESS, R. S. Phenomenal regression to the real object. *Brit. J. Psychol.,* 1931, **21**, 339-359; 1932, **22**, 1-30.

Department of Psychology
Princeton University
Princeton, New Jersey

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 155-160.

SOME POSSIBLE EXPLANATIONS OF SEX DIFFERENCES IN LANGUAGE DEVELOPMENT AND DISORDERS*

Department of Psychology, Fordham University

DOROTHEA McCARTHY

There has appeared in the literature considerable evidence indicating that American white boys are slightly later than girls in practically all aspects of language development which show developmental trends with age (3, pp. 551-555). These differences seldom are statistically significant, but the careful observer cannot ignore the amazing consistency with which these small differences appear in one investigation after another, each being conducted by a different experimenter, employing different techniques, different subjects, and sampling different geographical populations.[1]

Sex differences in favor of girls are present as soon as children begin to talk; that is, at about the age of onset of true language as distinct from prelinguistic utterances. Irwin and Chen (5) present curves showing the number of speech sounds from birth to 2½ years, which are practically identical for the two sexes for the first year of life, but these curves begin to diverge in favor of the girls in the second year of life when true speech emerges.

Although the developmental differences between the sexes are small in magnitude, they seem to be of considerable importance for the later acquisition of the more complex and secondary language forms for the effect seems to be cumulative. All statistics on the incidence of language disorders, particularly on the incidence of stuttering and of reading disabilities, reveal that language disturbances occur much more frequently among boys than among girls in most American reports (2, 6, 13). Series of case studies of language disorders show that rather than being evenly divided between the sexes, from 65 to 100 per cent of such disorders occur among boys.

The roots of the sex differentiation in language development must be sought in early infancy, for the differences appear at an extremely early age and the vital importance of imitative babbling in the establishment of language patterns is well recognized. Stengel (9) has stressed especially

---

*Received in the Editorial Office on September 13, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] A recent investigation by Anastasi and D'Angelo (*J. Genet. Psychol.,* 1952) reports a sex difference in favor of boys among five-year-old negroes.

the importance of the so-called "echo-reaction" stage in which the baby babbles back to the mother approximations of the sounds made by her. This is pleasant and satisfying to the infant and is considered to facilitate identification with the adult who provides the language model. Wyatt (12) also emphasizes the importance of the emotional quality of the early mother-child relations and the fact that the learning of the mother's speech is achieved through a process of unconscious identification. The mother is thus the child's first language teacher and as such is the first mediator of this vital cultural heritage.

The usual environmental situation in our culture is somewhat different for the boy infant than for the girl infant. Children of both sexes are usually cared for by, and have the constant companionship and speech model of, the mother. This experience in early language development is likely to be more satisfying for the girl than for the boy infant as she identifies more readily with the mother. The boy baby, on the other hand, needs to identify with, and to imitate the father's speech, of which he is likely to hear a minimum in our culture. When he does hear the father's speech and tries to imitate it, the experience must be much less satisfying for him than for the girl who can produce a fairly satisfactory echo-reaction to the mother's voice. The tremendous difference in voice quality of the adult male whose voice has changed with the attainment of maturity, and the small high-pitched voice of the boy infant certainly must make the echo-reaction stage much less satisfying for the male infant. It may even be quite confusing and perhaps even fear-producing, since loud sounds often produce fear in the young child. It may be argued that girls also experience the father's voice, about as much as do boys, but they probably feel less need to imitate and to identify with him than do boys for they are already making good progress in echoing the speech of the mother, and are finding considerable security and satisfaction in so doing. It is conceivable therefore, if this interpretation is correct, that contacts with the father would be less disturbing and less confusing to girl than to boy infants.

The writer has elsewhere (7 and 8) pointed out that language disorders in older children seem to be due to emotional insecurity based on attitudes of parental rejection. A second hypothesis which could account at least in part for the observed sex difference in language is that it may be due to differential parental attitudes towards the two sexes. Adoption agencies in the American culture usually have many more requests for girls than for boys for placement. If the same sex preference is present among true parents, that seems evident in instances in which prospective parents have their choice, it is reasonable to assume that girl babies are more often welcomed, and hence are given greater warmth, affection, and security from the very beginning, than are boys in our present day society.[2]

One of the reasons which may account for this preference is that boys are usually more active physically, and hence more difficult to control. Goodenough (4) has shown that boys have more frequent and longer anger outbursts than girls, that they are the objects of more disciplinary measures, and are usually handled more harshly than are girls. This would seem to indicate that they are more often frustrated by parental treatment, and must therefore experience a greater degree of emotional insecurity, which could lay the groundwork for later language disabilities as well as other behavior disorders.

If, as was pointed out above, the boy child finds less satisfaction in echoing the mother's speech, he will presumably repeat the activity less often than the girl, who as pointed out above, finds such activity more pleasant. Thus, there may be a differential in amount of linguistic practice, and in the degree of satisfaction derived therefrom by the two sexes in our culture which may tend to make a cumulative difference in experience in favor of girls.

In our culture, the boy is encouraged in active games, and even at the nursery age, when he is acquiring language skill, he is sent outdoors to play more often than the girl. This separates him from further adult linguistic stimulation. Boys are also more likely to engage in play with blocks and wheel toys and objects low in conversational value. Girls on the other hand are encouraged in indoor play with dolls, household toys, and table-play which have been shown to be of high conversation value (Van Alstyne, 10). Girls too are more likely to be permitted around the kitchen and other centers of household activity, and thus to have more constant adult attention. Girls are then more likely to be found within question-asking range of the mother and thus are more likely than their brothers to enjoy maternal contact and linguistic stimulation.

As the child finds satisfaction in echoing back the mother's utterances, the mother in turn finds it satisfying to engage in vocal play with the child. The girl child, however, being better able to identify with the mother, will probably be more responsive than the boy, so that the mother again may unconsciously give more linguistic practice to the girl than she does to the boy. Thus, it is conceivable, if these hypotheses are correct, that the groundwork is laid for greater conversation and companionship between mother and daughter than between mother and son. Such factors certainly seem worth considering in attempting to interpret the dynamics of the sex differences which are found. It becomes more understandable, in the light of such factors, why boys often interpret maternal behavior as rejecting and hence more of them feel insecure in their family relationship than is the case among girls.

---

[2] Studies of minority groups representing cultures which welcome boys should throw further light on this problem.

There are of course some instances of language disorders among girls. One may ask in the light of the above interpretations, how the relatively rare occurrence of language disabilities among girls can be accounted for. In the writer's clinical experience girls with language disorders have always had severely disturbed relationships with the mother so that the identification with the mother through language probably never occurred or was seriously distorted. Usually the relationships with the father were either non-existent or contacts were so infrequent or of such poor quality that they in no way compensated for the deprivation of maternal care, affection, and opportunity for identification.

Many language disorders are first noticed at the time of school entrance. Retardation and poor articulation which have been tolerated within the family circle become obvious. Unfavorable comparisons with other children contribute still further to the insecurity of already insecure children. At this time, adjustment must be made, not only to a new group of one's peers, often in a highly competitive situation, but also to another adult, usually another female, who becomes identified in the child's mind with the mother. When the mother-child relations in the preschool period have been of a wholesome quality, and a secure child enters the school situation, little difficulty arises. However, as has been outlined above, the average boy is more likely to enter school feeling somewhat insecure and rejected and to have had a less healthy relationship with his mother than the average girl. His predispositions therefore toward the female teacher are more likely to be fraught with a greater degree of anxiety, which may interfere with effective learning; and hence he is somewhat more likely to have difficulty with reading and other verbal skills in the academic situation than the average girl. Being about six months less mature linguistically than the average girl with whom he now has to compete in oral language, he is slower in beginning reading. His performance with the primer is likely therefore to be frustrating to the teacher whose aim is often centered on teaching reading skills early, and her behavior toward the boy is therefore likely to be one of further rejection, contributing still more to his insecurity. The basic anxiety so accumulated interferes with effective learning and the stage is set for the problem or typical boy behavior which Wickman (11) found so frustrating to teachers. If the child happens to encounter a teacher who is not well adjusted, who lacks insight into his behavior and into her own attitude, she will behave in what Anderson (1) describes as a "dominative" rather than an "integrative" way, and the vicious circle of frustration and aggression in the school situation becomes well established, and another non-learner in verbal skills, a potential problem child, is added to the rolls. It should also be remembered that the physical limitations of the classroom are undoubtedly more frustrating to the naturally more active boy than they are to the girl who is accustomed to and accepts sedentary play.

It is always possible to postulate differences in native endowment to account for the sex differences observed. Yet, it does not seem as though this explanation should be resorted to until account has been taken of the various environmental factors which certainly are operative in present day American culture. These interpretations are offered as tentative explanations which may operate as indicated in the long run when averages for large numbers of cases are considered. They are advanced in an effort to harmonize evidence from genetic studies on the one hand, and from various clinical investigations on the other hand. They must not be construed as operating in every individual case, for the individual differences within each sex group are indeed great, and hence almost everyone can point to individual cases of boys who are accelerated and girls who are retarded for their ages in language. It seems to the writer, however, that such dynamisms as have been outlined in this paper probably operate in the directions indicated with sufficient frequency to account for the small differences found in the averages for the two sexes at early ages in the genetic studies, and for the resultant cumulative effects observed in the clinical data at later ages.

REFERENCES

1. ANDERSON, H. H. The measurement of domination and of socially integrative behavior in teacher's contacts with children. *Child Devel.*, 1939, **10**, 73-89.
2. BENNETT, C. C. An inquiry into the genesis of poor reading. *Teachers College Contribution to Education, No. 442*, 1938.
3. CARMICHAEL, L. (*Ed.*). Manual of Child Psychology. New York: Wiley, 1946.
4. GOODENOUGH, F. L. Anger in young children. *Institute Child Welfare Monog., Ser.* 9. Minneapolis: Univ. Minnesota Press, 1931. Pp. xiii & 278.
5. IRWIN, O. C., & CHEN, H. Development of speech during infancy: Curve of phonemic types. *J. Exper. Psychol.*, 1946, **36**, 431-436.

6. LOUTITT, C. M. Clinical Psychology. New York: Harper, 1947. Pp. xviii & 661.

7. MCCARTHY, D. The psychologist looks at the teaching of English. *Indepen. Sch. Bull.*, 1946-47, No. 5.

8. ———. Personality and learning. *Amer. Coun. Educ. Stud.*, 1949, **35**, 93-96.

9. STENGEL, E. A clinical and psychological study of echo-reactions. *J. Men. Sci.*, 1947, **93**, 598-612.

10. VAN ALSTYNE, D. Play Behavior and Choice of Play Materials of Preschool Children. Chicago: Univ. Chicago Press, 1932.

11. WICKMAN, E. K. Children's Behavior and Teacher's Attitudes. New York: Commonwealth Fund, 1928.

12. WYATT, G. L. Stammering and language learning in early childhood. *J. Abn. & Soc. Psychol.*, 1949, **44**, 75-84.

13. YEDINACK, J. G. A study of the linguistic functioning of children with articulation and reading disabilities. *J. Genet. Psychol.*, 1949, **74**, 23-59.

Department of Psychology
Fordham University
New York City 58

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 161-168.

SEX DIFFERENCES IN ATTITUDES ABOUT SEX DIFFERENCES*

Department of Psychology, University of California

A. C. SHERRIFFS AND R. F. JARRETT

There is surprising agreement between men and women both with respect to the behaviors and characteristics which they attribute to males and females and to the values they place on these qualities. Moreover, there are remarkably few behaviors and attributes which are not uniformly ascribed by *both* men and women to one or the other of the two sexes. It would seem that virtually no behavior or quality escapes inclusion in either a male or female "stereotype," and that these stereotypes are substantially the same whether held by men or women.

Any right-minded reader is now objecting that we have generalized beyond whatever data we might have collected! What is the nature of these data?

While investigating certain aspects of the process of attitude change, the writers found it necessary to examine in some detail the characteristics ascribed by groups of men and by groups of women to men and women in general.[1]

A large number of behaviors and attributes which generally have value judgments applied to them were gathered from a wide variety of sources. Fifty-eight of these were selected for inclusion in a battery of items. Seventeen of these were judged appropriate to the cultural stereotype for males, 17 appropriate to the cultural stereotype for females. The remaining 24 items were believed not to be related to conventional stereotypes of masculinity or femininity. Half of the 58 items were judged by the authors and several colleagues to be positively valued in varying degrees and half to be negatively valued. This balance was maintained for both "stereotype" and "non-stereotype" items. The subjects were required, in a forced-choice situation, to state whether each item most appropriately characterized men or women.

It was thought desirable to conceal the intention to learn about the atti-

---

*Received in the Editorial Office on September 15, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] For complete discussion of the methods used in obtaining these data, as well as details concerning the context within which they were collected, see (1).

tudes of the subjects toward sex-differences, and an elaborate disguise was worked out. The details of this disguise are reported in (1), but one aspect of it which is of interest to readers of the present report was the requirement that the subject indicate whether each behavior or attribute most appropriately characterized not only one or the other of the two sex-groups, but also one or another of two age groups *and* one or another of three religious groups. (For part of our subjects this third set of categories was "only children" versus "people with sibs.")

On another occasion the subjects indicated the importance which they attached to each of the behaviors or attributes on the check-list by rating each item according to the following instructions:

> It is now important to know what this group feels as to the positive, neutral, or negative value of the following items. Alongside each item below, place a number (+3, +2, +1, 0, —1, —2, —3) describing the value of that item in your own scale of values. The seven numbers have approximately the following meanings:
>
> +3 The most positive of all values.
> +2 Highly positive: either very good, right, praise worthy, valuable, toward the ideal, etc.
> +1 Positive.
> 0 Neutral: neither positive nor negative, or indifferent in value, or equally positive and negative, etc.
> —1 Negative.
> —2 Highly negative: either very bad, blameworthy, wrong, unworthy, against the ideal, etc.
> —3 The most negative of all values.

The check-list included the following items (Table 1). They are pre-

TABLE 1
ITEMS OF CHECK-LIST

1. More people report their greatest happiness was given them by: (25)
2. Most faithful in marriage are: (17)
3. Trustworthiness has been found to be highest in: (50)
4. The persons who remain most loyal to friends and causes under fire are: (44)
5. Showing greatest emotional balance in crises are: (19)
6. Basic moral character is found to represent the highest values generally in: (47)
7. Most able to base actions on objective facts rather than on irrational personal feeling are: (34)
8. Having the most understanding of the real needs of children are: (2)
9. Possessing the keenest senses of humor are: (32)
10. Showing greatest warmth toward those less fortunate than themselves are: (14)
11. Least likely to get upset over small details are: (10)
12. The members of social service organizations who make the greatest individual contributions of time, effort, and results are: (26)
13. More courageous in the face of social disapproval are: (15)

TABLE 1 (*continued*)

14. Most conscientious in preparing for going to the polls on election day are: (9)
15. Most creative are: (40)
16. More courageous in the face of physical danger are: (37)
17. Most prone to punctuality out of consideration for others are: (38)
18. Impetus for future social progress will come to the greatest degree from: (55)
19. Most likely to be sensitive to the feelings of others are: (56)
20. More courageous in the face of pain are: (41)
21. Most insightful are: (16)
22. Showing most consideration for persons older than themselves are: (7)
23. General intelligence is highest in: (43)
24. The best leaders of small groups—5 to 10 individuals—more often are: (42)
25. Best losers at card games are: (8)
26. Least likely to exaggerate their personal difficulties are: (48)
27. Most imaginative are: (35)
28. Best equipped for the medical profession, other things being equal are: (11)
29. Least likely to have an opinion of themselves which is higher than the opinion of them by their peers are: (21)
30. More often tending to be competitive even in the most coöperative situations are: (33)
31. Physical defects are more common among: (46)
32. Personal conscience fails to inhibit reprehensible behavior most frequently in: (39)
33. Reasoning ability is poorest in: (18)
34. More likely to find satisfaction in daydreams than in actual life situations are: (51)
35. Fundamentally more prone to impose one's will on others are: (36)
36. Having least poise in awkward situations are: (53)
37. Predisposition to insanity is more frequent among: (5)
38. Least willing to accept responsibility for actions which will affect the happiness of others are: (23)
39. "Vanity" is a term applicable to more members of: (57)
40. "Keeping up with the Joneses" is more often a preoccupation of: (58)
41. More inclined to exhibitionism are: (52)
42. The sexual codes of their own group are more often violated by: (54)
43. "Carriers" of disease are more often found among: (20)
44. Most often swayed against their best interests by newspaper editorials are: (45)
45. Most likely to make pretenses socially rather than to reveal their real attitudes are: (29)
46. More likely to hold a grudge are: (6)
47. Most likely to give up a principle in order to obtain social status are: (22)
48. Poorest losers at sports are: (49)
49. Proportional to their numbers, most responsible for auto accidents in 1939 were: (1)
50. Life is *endured* more than enjoyed more often by: (12)
51. Prone to defend stubbornly their own point of view against overwhelming evidence that it is incorrect are: (24)
52. Most likely to be careless in matters of personal cleanliness are: (3)
53. Most subject to violent outbursts of temper are: (4)
54. If in charge of finances, most likely to go dangerously in debt are: (28)
55. Prone to take advantage of any situation to further their own selfish ends are: (30)
56. Individuals who are so dependent upon others that they have difficulty adjusting to life situations are more often: (31)
57. Most often likely to resort to unfair tactics in competitive situations are: (27)
58. More marital difficulties arise from the jealousy of persons who are: (13)

TABLE 2

VALUES AND SEX-GROUP DESIGNATIONS FOR THE FIFTY-EIGHT CHECK-LIST ITEMS BY MEN AND WOMEN SUBJECTS

(A)	(B)	(C)	(D)	(E)[a]	(F)[b]	(G)[c]	(H)[d]	(A)	(B)	(C)	(D)	(E)[a]	(F)[b]	(G)[c]	(H)[d]
		Value								Value					
Item	True of men or of women	Men ($N = 179$)	Women ($N = 212$)	(t^1)	% men saying men ($N = 347$)	% women saying men ($N = 492$)	Difference (F)-(G)	Item	True of men or of women	Men ($N = 179$)	Women ($N = 212$)	(t^1)	% men saying men ($N = 347$)	% women saying men ($N = 492$)	Difference (F)-(G)
1	F	2.37	2.59	— .98	33.4	46.5	—13.1	30	M	— .48	— .98	1.36	70.0	58.7	11.3
2	F	2.34	2.59	—1.04	20.2	18.5	1.7	31	M	— .56	— .37	— .68	68.3	64.2	4.1
3	M	2.25	2.44	— .88	69.5	70.1	— .6	32	M	— .69	.70	.03	61.1	63.4	— 2.3
4	M	2.06	2.11	— .22	76.4	73.4	3.0	33	F	— .85	—1.48	1.54	8.6	22.0	—13.4
5	M	2.03	2.14	— .34	76.4	67.5	8.9	34	F	— .94	—1.36	1.48	21.3	9.3	12.0
6	F	1.98	2.43	—1.69	33.7	23.0	10.7	35	M	— .99	—1.31	.97	53.6	50.6	3.0
7	M	1.97	1.84	.52	87.3	87.4	— .1	36	M	—1.04	—1.14	.38	80.4	80.3	.1
8	F	1.93	2.08	— .64	7.2	3.0	4.2	37	F	—1.08	—1.02	— .18	47.8	38.0	9.8
9	M	1.73	1.73	.00	91.1	79.5	11.6	38	M	—1.17	—1.65	1.08	51.0	56.1	— 5.1
10	F	1.72	1.98	—1.16	28.5	20.9	7.6	39	F	—1.17	—1.29	.39	10.1	13.2	— 3.1
11	M	1.66	1.74	— .41	71.5	80.3	— 8.3	40	F	—1.25	—1.34	.31	5.5	3.9	1.6
12	F	1.59	1.72	— .52	21.6	10.6	11.0	41	M	—1.25	—1.48	.71	58.5	67.7	— 9.2
13	M	1.55	1.56	— .03	86.2	83.5	2.7	42	M	—1.27	—2.00	2.24	83.0	85.4	— 2.4
14	M	1.52	1.56	— .18	78.1	74.8	3.3	43	M	—1.30	—1.19	— .31	63.1	65.4	— 2.3
15	*	1.51	1.49	.08	66.6	46.7	*19.9*	44	F	—1.42	—1.35	— .28	21.6	21.1	.5
16	M	1.50	1.48	.11	85.9	82.9	3.0	45	F	—1.44	—1.66	.93	11.8	5.9	5.9
17	M	1.49	1.50	— .05	74.9	66.9	8.9	46	F	—1.51	—1.80	1.12	40.6	29.5	11.1
18	M	1.47	1.35	.34	69.5	50.4	19.1	47	F	—1.57	—1.80	.72	21.9	23.4	—-1.5
19	F	1.44	1.62	— .54	21.6	14.0	7.6	48	*	—1.57	—1.66	.29	51.6	42.1	*9.5*
20	*	1.42	1.51	— .40	53.9	40.4	*13.5*	49	M	—1.61	—1.95	.97	73.5	78.5	— 5.0
21	*	1.42	1.44	— .07	58.8	27.2	*31.0*	50	F	—1.64	—1.87	.69	27.1	33.3	— 6.2
22	F	1.41	1.86	—1.33	47.3	32.7	14.6	51	*	—1.72	—1.82	.24	47.5	54.5	— *7.0*
23	M	1.34	1.39	— .11	84.1	69.5	14.6	52	M	—1.79	—2.09	1.05	88.5	88.4	.1
24	M	1.27	1.16	.47	81.8	61.4	20.4	53	F	—1.83	—1.92	.30	47.5	35.4	12.1
25	M	1.25	1.37	— .62	68.9	69.3	— 1.6	54	F	—1.84	—2.05	.90	45.5	48.0	— 2.5
26	M	1.19	1.40	— .70	71.8	81.3	— 9.5	55	F	—1.88	—2.25	1.28	44.4	32.7	12.7
27	F	1.18	1.16	.08	30.8	12.2	18.6	56	F	—1.93	—2.06	.52	28.8	24.4	4.4
28	M	.98	— .18	2.09	95.4	92.7	2.7	57	M	—1.99	—2.26	1.16	57.9	57.9	.0
29	M	.38	.82	—1.18	64.8	60.0	4.8	58	F	—2.04	—2.28	.92	25.9	20.5	5.4

[a] A negative sign to this *t* means that the women assign this item a higher value (more positively) than do the men.

[b] The standard error of a proportion of 50 per cent for this sample is approximately 1.87 per cent. Thus the 1 per cent confidence interval is 45.2-54.8.

[c] The standard error of these proportions is about 1.45 per cent. Thus the 1 per cent confidence interval is 46.2 per cent-53.8 per cent.

[d] The standard error of the difference between these percentages depends, of course, upon the value *P* (the weighted mean proportion of the subjects who say the item characterizes men). When this *P* is close to 50 per cent, the standard error of the difference is about 1.1 per cent; when close to 70 per cent (or 30 per cent) it is about 1 per cent;

sented here in order of decreasing mean value as rated by our male subjects, though it will be noted in the table which follows that comparatively minor rearrangement would be required to order them in terms of the ratings of our women subjects. The numbers in parentheses following each item are the order-numbers of the items on the original list as it was presented to the subjects.

Table 2 summarizes our data. The first column identifies the item in terms of its number in the above list. The second column shows the sex-group characterized by the item according to our subjects. (Disagreement between men and women subjects on this matter is indicated by an asterisk in this column.) Columns 3 and 4 contain the mean value-ratings by men and women, respectively, and in Column 5 is found the *t* for testing the significance of the difference between these two means. (A negative *t* indicates that the women gave a more positive value to the item than did the men.) Columns 6 and 7 show the per cent of men and women subjects, respectively, asserting that the item characterizes men. (The second column, of course, depends upon these entries.) Column 8 presents the differences in the proportion of men and women attributing the item to men. The footnotes to this table set forth approximate standard errors for the several values tabled.

These data raise some interesting questions about the attitudes of men and women toward their own and the opposite sex. They also stimulate speculation concerning the origin, nature, and structure of stereotypes in our culture as well as concerning the learning and modification of social values by individuals.

It will be noted that there is remarkable agreement between the mean values assigned by men and women to the several behaviors and attributes in our list. At first we were inclined to assert that our data suggest that with respect to no item did the mean of the population of men differ from the mean of the population of women. The *t*'s reported in Column 5 of Table 2 would seem to justify this conclusion. In no case is the *t* significant at the 1 per cent level, and in only two cases (less than the 2.95 to be expected by chance) is that mean difference significant at the 5 per cent level. Closer scrutiny, however, reveals that our preliminary conclusion is untenable; of the 29 positively valued items (Items 1-29), the women assigned higher values than the men in 21, while for the negatively valued items, the women assigned lower values than the men in 25 out of 29 cases. Thus the women seem to place more extreme values on most of our behaviors and attributes, even though they agree very closely with the men as to the value which any given behavior deserves. The probability that our women would assign more extreme values to 46 or more of the 58 items if in fact men and women agree exactly about these values is considerably less than 0.001.

Our items cover a wide range of values and a wide variety of behaviors and attributes, and at the time they were selected it was not anticipated that men and women would agree so closely with respect to the values placed upon them. Indeed, had we been asked to make a prediction about the values which they would assign, we would have anticipated that each sex would place slightly higher values on those items seen as characterizing its own group. We might further have expected that women would have been more conservative—would have rated the items as less extreme than did men.

Reference to Table 2 also reveals that there are surprisingly few items which are not attributed by significantly more than 50 per cent of both men and women subjects to one or the other sex-group. Our women subjects are equivocal in only five of the 58 items—1, 15, 18, 35, and 54. Only for these items is the proportion of women who attribute the behavior to one or the other sex-group not significantly (at the 1 per cent level) different from 50 per cent. Our men fail in 9 of the 58 items to attribute the behaviors or attributes clearly (at the 1 per cent level) either to men or to women. For these subjects the equivocal items are 20, 22, 35, 37, 38, 48, 51, 53, and 54. This larger number of items not clearly attributed by men subjects to one or the other sex-group may be without significance, since the smaller sample of men has a larger standard error, and therefore we would expect more "insignificant" differences for this group, even should real differences exist for all items. It is interesting, but not statistically significant, that seven of the nine items which the men fail to assign convincingly to one or the other sex are items to which both men and women assign negative values (35, 37, 38, 48, 51, 53, and 54).

Since there are two items (35 and 54) common to both lists, it will be seen that of our 58 items there are only 12 in which we fail to find *both* men and women ascribing the behaviors and attributes to one of the sex-groups. In 10 of these 12 items only one of the groups is equivocal and in only *two* are both groups equivocal.

In addition to the fact that in the case of only a small number of items do we find either men or women doubtful as to the sex-group assignment of an item, we find only five items (15, 20, 21, 48, and 51) on which men and women do *not agree* as to the sex-designation. Four of these were items considered equivocal by either our men or our women subjects, but in each of these cases the difference between the proportions of the two groups of subjects attributing the item to men (or to women) was highly significant.

Keeping in mind the method by which our items were selected and the criteria for inclusion of items, the summary of our findings to this point may be stated in the words of our opening paragraph:

> There is surprising agreement between men and women both with respect to the behaviors and characteristics which they attribute to males and females and to the values they place on these qualities. Moreover, there are remarkably few behaviors and attributes which are not uniformly ascribed by both men and women to one or the other of the two sexes. It would seem that virtually no behavior or quality escapes inclusion in either a male or female "stereotype," and that these stereotypes are substantially the same whether held by men or by women.

We have not yet justified the last two sentences of this summary. Such justification must take the form of a contention either (*a*) that the behaviors and attributes which we included in our list can reasonably be considered a random sample of all behaviors and attributes which people use in the process of forming judgments about other people or in describing other people, or (*b*) that even though this may not be so, there are good reasons for believing that the results are not essentially different from those which would have been obtained with such a random sample. We propose to make this latter assertion.

We cannot claim that our behaviors comprise a random sample from the population described, since we purposely included 34 items which we thought were generally considered characteristic of either men or women. The remaining 24 items were selected *not* to conform to a masculine or feminine stereotype. Thus while we might consider that we have an approximation to a random sample of "stereotype" items and an approximation to a random sample of "a-stereotype" items, we have no reason to believe that we have included these two types of items in their proper proportions, whatever these may be.

However, the singular lack of success which accompanied our attempt to select a-stereotypic items convinces us that the proportion of such items in the population must be very small indeed, and that our results are substantially the same as those we would obtain with a random sample of behaviors and attributes.

Moreover, it is inconceivable to us that the process of learning which underlies the ability of our subjects so readily to make such judgments as we (and everyday life) required of them could possibly involve learning associations between sex-group and *each* of the conceivable behaviors and attributes which comprise together the population of items referred to above.

We would suggest that our subjects learn that there are a number, perhaps a relatively small number, of rather general traits which characterize the behavior of men (or women). This learning may be by way of direct personal experience with men and women (this is perhaps the most important route) though for other attitudes this might not be so, or by way of experience with the verbalized beliefs and attitudes of their associates. In either case, each of the behaviors or attributes on our list is related by association with one or more of these "general traits," and by way of this association also associated with either "men" or "women." It is this pattern of "general traits" which properly constitutes the "stereotype."

Economy of thought demands that in the continual learning process of existence the human organism assign every particular "event" (including such particular behaviors and attributes as, e.g., being responsible for automobile accidents, proneness to be swayed by newspaper editorials, etc.) to one or more of a relatively small number of cluster-like categories. Each of these categories possesses some integrity and stability, the various elements belonging to the category being associated with one another partly by virtue of their common belongingness.

Then when our subjects are asked to associate some particular behavior with a particular sex-group (or conversely), the associative process is a mediate one—particular behavior (or attribute) to "general trait," to sex-group. Similar mediate association might be expected were we to require judgments concerning other categories of persons, e.g., Negroes, "only children," etc., so that we would expect to find the same "remarkably small proportion" of such a series of items as ours not attributed to one or another of such groups.

Certainly this description is greatly oversimplified, and the reader may be tempted, as we are, to elaborate on it.

REFERENCE

1. JARRETT, R. F., & SHERRIFFS, A. C. Propaganda, debate, and impartial presentation as determiners of attitude change. *J. Abn. & Soc. Psychol.,* (In press).

Department of Psychology
University of California
Berkeley 4, California

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 169-179.

THE INTERCORRELATIONS BETWEEN MEASURES OF SCHOOL LEARNING*

Department of Education, Yale University

J. W. Tilton

A. The Problem

This article supplements an earlier one (7). In the earlier article protest was made against the use of poor measures of learning as criteria for testing the concept of intelligence as a general ability to learn. The measures of learning used have been very unreliable. Furthermore, no thought seems to have been given to the possibility that there was relatively little opportunity for the ablest to profit from the learning experience, nor to the possibility that the tests used could not register what the ablest did learn. With these considerations in mind, measures of learning were described in the earlier article which correlated highly enough with intelligence test scores to be consistent with the idea of intelligence as ability to learn.

At that time, no attempt was made to examine the generalness of learning. The present article describes such an examination of learnings in a school situation. The report consists of five parts: (*a*) a description of the data to be used as to their adequacy along two of the three lines mentioned above; (*b*) a description of some further requirements when the generalness of learning is in question; (*c*) presentation of data as to the generalness of learning in one school situation at the fourth and fifth grade level; (*d*) a comparison of the factor pattern at the fourth and fifth grade level with the factor pattern of prior gains; and (*e*) some observations concerning the predictability of improvements in score at the fourth and fifth grade level.

B. The Data

The tests used were the Unit Scales of Attainment. The initial testing was done with the Division I battery which is designed for Grades 3 and 4. Forms *A* and *B* were administered in the fall to the fourth grades used by the New Haven State Teachers College for teacher-preparation purposes. The end tests were Forms *A* and *B* of the Division II battery designed for Grades 5 and 6. They were administered in the spring about 20 months after the initial testing.

---

*Received in the Editorial Office on October 5, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

TABLE 1
RELIABILITIES

	Initial scores	Final scores	Improvements
Reading	.791	.856	.364
Social Studies	.838	.950	.694
Literature	.831	.834	.521
Science	.825	.768	.360
Arithmetic Problems	.835	.910	.629
Arithmetic Fundamentals	.852	.907	.769
Spelling	.964	.934	.692
English	.889	.893	.744
Averages	.85	.88	.60

Two features of this procedure were designed to give better reliabilities than are ordinarily obtained, two-year gains and two-form measures. The reliabilities shown in Table 1 are Form *A* with Form *B* coefficients stepped up with the Spearman-Brown formula for the Form *A* and Form *B* combined measures.

Were the tests valid enough to provide full measures of the learning of all pupils?

It might be claimed that no standardized achievement test battery measures the outcomes of enriched teaching for the ablest students. That may be the case. But so far as standardized achievement testing can take care of the situation, there was plenty of opportunity in the end tests for the abler students to make high scores. The Division II test range of difficulty did not impose any obvious restrictions upon the measures of improvement.

C. SPECIALIZING INFLUENCES

The measures are fairly reliable. There is plenty of "top" in all the tests. If general improvement is not found, can a general learning ability be denied? By no means. Specializing influences could have counteracted the influence of the general learning ability.

Suppose the group were composed of upper-class college students with a variety of major interests and goals, pre-medical, pre-engineering, etc. The classics majors would make the largest gains in language, the lowest gains in science and mathematics. Would low or even negative correlations between gains mean that there is no general ability to learn? How far up the educational ladder do pupils go without specialization of interest and application? Possibly specialization is present from the start, gradually reducing the evidence for a general ability. If that is the case, the fourth and fifth grades are a good level at which to secure data; a compromise between the lower level at which so few "subjects" are taught and the upper level at which greater specialization is to be expected.

There are, however, other specializing influences. There are specializing influences in the teaching pattern. It was pointed out in the earlier article that teaching may provide plenty of opportunity for all to learn or it may cause the ablest pupils to mark time while those of lesser ability are helped to make large gains. The only safeguard against this, in the present case, is the probability that the teachers were better than average teachers, and that more than usually is the case, each child was given opportunity and help according to his ability.

But can it be assumed that the 10 teachers involved were alike in their ideas of the relative importance of the eight areas measured, alike in their relative interest in these areas, and alike in their relative abilities to teach the areas effectively? Perhaps the largest gains in reading were to be had only by being in the class taught by Miss *X*, and perhaps the smallest gains in arithmetic were made by the pupils taught by Miss *Y*. Under such circumstances, lack of intercorrelation between reading and arithmetic in the combined group would but reflect the heterogeneity of the teaching patterns and would be of little worth as evidence against the existence of a general ability to learn.

In order to eliminate such heterogeneity, one should collect data for many classes so that classes could be dropped freely without serious loss of reliability in the resulting homogeneous group. In the present case, data were available for only eight different instructional groups. Very little elimination was possible. Only two groups were eliminated, and this was done crudely by inspection. They were the two groups which exhibited the most unique pattern of gains. Group *A* made obviously large gains in social studies, literature, science, and arithmetic, but made low gains in English. Group *B*, on the other hand, made high gains in English but low gains in arithmetic fundamentals. The intercorrelations for the total group ($N=162$) and for the reduced group ($N=102$) after Groups *A* and *B* were eliminated are shown in the first two columns of Table 2. As was to be expected, the correlations between social studies, literature and science were decreased by the elimination of the two groups. The effect upon the correlation between arithmetic problems and arithmetic fundamentals could not be predicted because the elimination of Group *A* would tend to lower it while the elimination of Group *B* would tend to raise it. The correlations of gains in these five areas with gains in English were raised as was expected. The over-all average change was slight but the intercorrelations for $N=102$ are sub-

TABLE 2
INTERCORRELATIONS BETWEEN GAINS

	$N=162$	$N=102$	Boys $N=48$	Girls $N=54$	Boys and girls *r*'s averaged	$N=102$ corrected for attenuation
R with *So*	.216	.187	.225	.209	.217	.372
L	.143	.154	.036	.228	.132	.354
Sc	.234	.221	.240	.202	.221	.609
AP	.230	.316	.332	.314	.323	.660
AF	.210	.266	.304	.237	.270	.502
Sp	.073	.065	.145	.021	.083	.129
E	.227	.237	.317	.173	.245	.456
So with *L*	.699	.590	.580	.744	.662	.981
Sc	.676	.314	.368	.334	.351	.629
AP	.407	.430	.441	.459	.450	.651
AF	.369	.408	.399	.480	.439	.564
Sp	.258	.164	.253	.125	.189	.237
E	.311	.400	.406	.411	.409	.556
L with *Sc*	.607	.233	.254	.239	.246	.538
AP	.361	.319	.265	.393	.329	.557
AF	.406	.426	.279	.525	.402	.672
Sp	.194	.171	.122	.194	.158	.285
E	.192	.334	.293	.386	.339	.537
Sc with *AP*	.383	.325	.537	.115	.326	.683
AF	.286	.142	.244	.066	.155	.270
Sp	.202	.171	.150	.186	.168	.342
E	.121	.268	.320	.234	.277	.518
AP with *AF*	.478	.536	.504	.590	.547	.770
Sp	.213	.251	.303	.236	.270	.381
E	.255	.366	.375	.354	.365	.535
AF with *Sp*	.303	.227	.123	.292	.208	.311
E	.380	.527	.481	.579	.530	.697
Sp with *E*	.389	.345	.205	.462	.334	.481
Averages	.315	.300	.304	.314	.309	.510

mitted as a more valid reflection of ability to learn because of *some* elimination of what is interpreted as heterogeneity of teaching pattern.

In Columns 3 and 4, the intercorrelations are shown for boys and for girls. They are shown separately for two reasons. One is to show that in spite of the reduction from 162 to 102, there is enough reliability left in the data and enough positive intercorrelation to have prevented the occurrence of a single negative correlation. The other reason for the separation was to test the influence of the specializing influence of sex upon the intercorrelations reported in Column 2. It was predicted that the average *r*'s in Column 5 would tend to be larger than the *r*'s in the second column. They did tend to be larger, and the average difference was 3.2 times its standard error.

Other things being equal, intercorrelations for boys or for girls would be better data for the present purpose; but because of the greater reliability of the *r*'s in Column 2 for boys *and* girls, the Column 2 *r*'s are submitted as the best data available under the circumstances. They are shown in the last column of Table 2 corrected for attenuation.

D. EVIDENCE OF A GENERAL ABILITY TO LEARN

The absence of negative coefficients in Table 2 is itself evidence of the generality of learning in the eight areas. The substantial loadings with Factor 1 shown in Table 3 are as good evidence of a general ability to learn as could be expected. In view of the many things which influence progress in addition to ability, the data may be claimed to be quite consistent with a concept of general ability to learn. They do not prove its existence any more than a failure to find a general factor[1] would prove that there is no general ability to learn. *Without* a general ability, circumstances could combine to produce a general factor; *with* a general ability they could combine to obliterate the evidence of its existence.

TABLE 3
FACTOR LOADINGS BEFORE ROTATION FOR THE EIGHT MEASURES OF IMPROVEMENT, BASED ON THE CORRECTED *r*'S IN THE LAST COLUMN OF TABLE 2

Area	Factor 1	Factor 2	Factor 3	Factor 4	h^2
R	+.63	−.18	+.46	−.10	.65
So	+.84	+.52	−.09	−.07	.99
L	+.83	+.46	−.28	−.15	1.00
Sc	+.72	+.08	+.37	+.20	.71
AP	+.85	−.14	+.18	−.12	.79
AF	+.77	−.33	−.24	−.41	.94
Sp	+.45	−.20	−.18	+.43	.46
E	+.76	−.20	−.22	+.17	.70

For these data, it is thought that the loadings shown in Table 3 are more meaningful than those required for Thurstone's simple structure (5). However, for those who do not share this opinion (2), loadings after three rotations, following the Peters and Van Voorhis algebraic procedure (4), are shown in Table 4. As may be seen in Column 1, Factor 1 remains a general factor in spite of the modifications in it which were required by the transformations of Factors 2 to 4.

E. THE PATTERN OF 4-5 GAINS COMPARED WITH THE PATTERN OF PRIOR GAINS

The improvements in score made during the fourth and fifth grade experiences are referred to as 4-5 gains. The initial scores used in obtaining them

---

[1] Factor 1 in Table 3 is Burt's General Factor (1).

TABLE 4
LOADINGS AFTER ROTATION

Area	Factor 1	Factor 2	Factor 3	Factor 4	h^2
R	.39	.09	.70	.08	.66
So	.52	.81	.16	.18	.99
L	.64	.76	.00	.15	1.00
Sc	.28	.37	.61	.35	.70
AP	.65	.22	.53	.20	.79
AF	.95	.01	.15	.00	.92
Sp	.28	.00	.05	.62	.46
E	.64	.12	.15	.50	.70

are, it is maintained, measures of gain up to the beginning of the fourth grade. They are referred to as 0-3 gains. They are the gains made from zero knowledge and ability during the pre-school years and the three years in Grades 1 through 3. To what extent does the pattern of 4-5 gains resemble the pattern of 0-3 gains?

The 0-3 gains with the exception of arithmetic fundamentals were more general than the 4-5 gains. This may be seen in Table 5, (*a*) in the larger loadings with the general factor (Factor 1), (*b*) in the larger communalities, and (*c*) in the fact that the 0-3 intercorrelations are described with three factors where four were required for the 4-5 intercorrelations. There was no doubt about three factors being sufficient. All 36 of the final residuals were within .033 of being equal to zero.

TABLE 5
COMPARISON OF FACTOR LOADINGS AT THE 4-5 AND PRECEDING LEVELS

	Factor 1		Factor 2		Factor 3		Factor 4	
	0-3	4-5	0-3	4-5	0-3	4-5	0-3	4-5
R	.96	.63	.19	—.18	.15	.46		—.10
So	.91	.84	.34	.52	—.06	—.09		—.07
L	.94	.83	.33	.46	—.09	—.28		—.15
Sc	.96	.72	.26	.08	—.01	.37		.20
AP	.89	.85	—.33	—.14	—.24	.18		—.12
AF	.71	.77	—.43	—.33	—.42	—.24		—.41
Sp	.78	.45	—.15	—.20	.44	—.18		.43
E	.83	.76	—.20	—.20	.22	—.22		.17

Factor 2 loadings show similar though not identical patterns. The factor is the same in showing the arithmetic and language skills to be grouped in opposition to the information subjects, social studies (history and geography), literature, and science. The factor is different in that reading was associated with the skill subjects at the 4-5 level but was associated with the information subjects at the 0-3 level.

The signs must be both plus and minus for Burt factors subsequent to the first. Which are plus and which are minus is immaterial. The facts could be presented as satisfactorily with all signs changed. What is revealed is this. If each pupil's gains were portrayed graphically as a profile around his own average gain, it would be seen that the information gains would *tend* to be all crests, or all troughs. That is, if one were high (or low) in the profile, the others would tend to be high (or low). The same could be said of the skill gains. Also, it would be seen that for the profiles in which the crests were provided by the information subjects the troughs would be provided by the skill subjects and vice versa. In other words, on the basis of plus and minus loadings for Factor 2, one should not infer anything as to the likelihood of an area being high or low in the profile. The signs tell rather which ones tend to be together on the same side of the base line and which opposite.

What is the meaning of the difference with regard to reading? The first thing to note is that no one column tells the whole story about reading (or any subject). It would be a mistake to interpret the Factor 2 loading for reading as indicating a negative relationship between progress in ability to read and the amount of information possessed in the social studies, literature, and science areas. Such an inference would be contrary to the correlation shown in Table 2 and to the loadings with Factor 1. What it means is that the reading gains are less closely associated with general progress during the 4-5 years than they were during the 0-3 period. This could be due to the acquisition by the less able pupils during the 4-5 period of a reading skill which was acquired in the earlier period by the more able. It might be impossible for the ablest pupils to continue to improve in ability to read at the rate at which they progressed during the earlier period. At any rate, 4-5 progress in social studies (history, geography) and literature was more closely related to 0-3 gains in reading than to the 4-5 gains in reading. Progress in these information areas during the fourth and fifth grades seemed more a function of ability to read at the beginning of that period than of the gains in reading during the period.

In this difference between the 0-3 and 4-5 patterns, science gains (and to a still lower degree arithmetic problems) were associated with reading gains in being less related to general progress. Progress in arithmetic fundamentals at the 4-5 level was *more* highly correlated with general progress than it was at the 0-3 level. Thus it appears that gains in a subject-matter area may vary from level to level as to their general significance. Progress in reading, science, and arithmetic fundamentals at the 0-3 level showed no correlation with progress at the 4-5 level. The corrected *r*'s were .00, —.03, and —.03 in these respective areas. On the other hand, for social studies, literature, arithmetic problems, spelling, and English the *r*'s were .72, .42, .26, .32, and .60.[2]

TABLE 6

PROGRESS IN READING AS RELATED TO 4-5 PROGRESS IN SOCIAL STUDIES AND LITERATURE

	Reading 0-3 Gains		Reading 4-5 Gains	
	Uncorrected *r*'s	Corrected *r*'s	Uncorrected *r*'s	Corrected *r*'s
Social Studies 4-5 Gains	.48	.64	.19	.37
Literature 4-5 Gains	.32	.50	.15	.35

This difference in relationship was so influential as to have determined the nature of the third 4-5 factor. It was only after Factor 3 accounted for this difference that the analysis revealed another similarity between the 0-3 pattern and the 4-5 pattern. The 0-3 Factor 3 and the 4-5 Factor 4 both show an association between the gains in the two parts of arithmetic, and also between the gains in spelling and the remaining parts of language usage included under the head of English. Most prominently they agree in showing opposition between arithmetic fundamentals and spelling.

When it is kept in mind that the 0-3 gains cover both a longer period and an earlier one, perhaps the patterns of intercorrelation are as similar as they should be expected to be. If they are, what ground is there for mistrusting end scores as measures of learning? For instance, for the data being described, do not (*a*) the initial scores, (*b*) the differences between initial and final scores, and (*c*) the final scores all have the same kind of meaning and validity? The initial score is a measure of learning up to the beginning of Grade 4. The difference between final and initial scores is a measure of learning during the fourth and fifth grades, and the final score is a measure of learning up to the end of Grade 5. *And in general, do not the conclusions about ability to learn drawn over the years from end scores have just as much validity as if they had been drawn from final minus initial scores?*

F. PREDICTABILITY OF GENERAL PROGRESS AT THE 4-5 LEVEL

What implications are there in the data for the practice of those who could use in advance some index of the 4-5 gains in score? There are no direct implications concerning the use of intelligence test scores for this purpose. Intelligence test scores were not obtained. It was thought that a request for more time than was needed for the two-form achievement-testing program would have been excessive. It was also thought that if a general learning factor were discovered, substantial correlation between the general factor and intelligence test scores could be assumed, because the high communality between intelligence and achievement scores (3) is accepted without question. The chief question has been whether there is a general learning factor.

---

[2] These *r*'s were based upon two uncorrected *r*'s, the 0-3 Form *A* with the 4-5 Form *B* and the 0-3 Form *B* with the 4-5 Form *A*. They are thus not diminished by the negative correlation of errors which is ordinarily involved in the correlation of initial scores with gains.

It *can* be reported that the general learning factor (Factor 1) described in Table 3 is correlated —.52 with chronological age. Or, to phrase it more precisely, when chronological age and an educational quotient based on initial and final scores combined were included in the matrix along with the eight measures of learning, the Factor 1 correlation was —.52 with chronological age and was .72 with educational quotient. The average uncorrected *r* of chronological age with the separate gains was —.24 and it was —.25 with the initial scores from which the gains were measured.

When an unweighted average of the 4-5 gains is used as the measure of general progress in the fourth and fifth grades, the possibilities for predicting it from age and achievement test data are shown in Table 7. The *r*'s are not high but every one of the 11 measures is an index of subsequent general improvement. The science measure alone was a better intelligence test than the whole battery and age as they are combined in *EQ*. Arithmetic fundamentals was the poorest index.

TABLE 7

THE POSSIBILITIES FOR THE PREDICTION OF GENERAL PROGRESS IN GRADES 4 AND 5 FROM MEASURES OF PRIOR PROGRESS

	Uncorrected *r** with criterion	Corrected *r* with criterion
R	.38	.61
So	.34	.52
L	.32	.49
Sc	.45	.71
AP	.31	.47
AF	.15	.22
Sp	.29	.39
E	.37	.55
EA	.36	.50
CA	—.31	—.40
EQ	.41	.56

*The *r*'s in this column except for *CA* are averages of two coefficients Form *A* with Form *B* criterion and Form *B* with Form *A* criterion. The *CA r* is the average of the *r* between *CA* and Form *A* criterion and the *r* between *CA* and Form *B* criterion.

G. SUMMARY AND CONCLUSION

"Scholastic ability is general. Get a good measure of a pupil's ability to learn, and you may expect him to measure up to it in all the major subjects." This line of thinking has been used very commonly and confidently at all scholastic levels. Only in recent years has a question been raised as to the validity of a concept of a general ability to learn. The questioning is desirable and profitable (6). The present article and the earlier one referred to above were not designed as a protest against critical examination. Protest is directed solely to the use of poor and inappropriate measures. Considerable space was therefore given above to the discussion of factors which might diminish or even prevent positive intercorrelation between measures of school learning. Data were secured and selected in which the prospects were thought to be fairly good for a general ability to manifest itself, if there is one. It was pointed out that a generalness of progress does not constitute *proof* of a general ability to learn, any more than lack of agreement constitutes disproof.

The intercorrelations between the measured gains were all positive and substantial. The Burt first or general factor remained a general factor when the three bipolar factors were rotated to have zero to positive loadings. The gains in score were not as general as were the initial scores, themselves measures of what had been learned up to the time of the initial test. Less generality at an upper level is not inconsistent with the idea of a general ability to learn, for a general ability is not inconsistent with an increasing specialization.

The pattern of intercorrelation at the two levels, during the fourth and fifth grades and before, was similar. This was in spite of the facts that the periods differed in length, that one was partly pre-school, and that the measures were end-tests at one level and differences between initial and final scores at the other level. Such differences as appeared are interpreted as examples of ways in which the teaching situation may prevent a general ability to learn from manifesting its influence. It seems, for instance, that to a certain extent the fourth and fifth grades were a period for the less able pupils to catch up in their ability to read while the more able pupils used the ability they had already acquired. At any rate, progress at the fourth and fifth grade levels in three areas, reading, science, and the fundamental operations of arithmetic, was uncorrelated with progress in these areas prior to the fourth grade. With these three exceptions, the tendency was for greater progress in the fourth and fifth grades to be made by those who had made the greater prior progress.

On the whole, the data presented here, and in the earlier article, are fairly consistent with a concept of a general ability to learn and with the identification of it with the "general" intelligence test.

REFERENCES

1. BURT, C. The Factors of the Mind. New York: Macmillan, 1941.
2. CATTELL, R. B. Factor Analysis. New York: Harper, 1952.
3. KELLEY, T. L. Interpretation of Educational Measurements. Yonkers-on-Hudson: World Book, 1927.
4. PETERS, C. C., & VAN VOORHIS, W. R. Statistical Procedures and Their Mathematical Bases. New York: McGraw-Hill, 1940.
5. THURSTONE, L. L. Multiple Factor Analysis. Chicago: Univ. Chicago Press, 1947.
6. TILTON, J. W. An Educational Psychology of Learning. New York: Macmillan. 1951.
7. ———. Intelligence test scores as indicative of ability to learn. *Educ. & Psychol. Meas.*, 1949, **9**, 291-296.

Department of Education
Yale University
28 Hillhouse Avenue
New Haven, Connecticut

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 181-195.

MANUAL FOR SCORING THE THEMATIC APPERCEPTION TEST*[1]

VA Regional Office, Wilkes-Barre, Pennsylvania

FRANKLYN D. FRY

A. OBJECTIVE

The objective of this scoring scheme is to make the scoring of the *TAT*, with reference to certain tendencies or trends, as effective and efficient but yet as simple as possible, one which would be diagnostic and differentiate one group from another by indicating the tendency or trend of the subjects of one group to attribute to or to project onto the characters in their protocol stories certain types of moods, attitudes, personality traits, and emotional states or qualities which the subjects of the other group did not attribute to or project onto the characters in their protocol stories. There is no one test or scoring scheme which will reveal all of the various personality traits, qualities, and characteristics which a subject may possess; however, a multiphasic personality evaluating instrument like the Bernreuter Personality Inventory, the Minnesota Multiphasic Personality Inventory, the Rosenzweig Picture-Frustration Study, and this scoring scheme for the *TAT* can all make a contribution towards gaining insight into the personality which an individual may possess.

It should be noted that the *TAT* Scoring Blank, which the author has developed, may be used to score just one protocol story, or it may be used to summarize all the scores of the 20 individual protocol stories. Since check marks are used on the *TAT* Scoring Blank, when evaluating individual *TAT* protocol stories, to indicate the predominating characteristics and qualities; and since tally marks are used on the *TAT* Scoring Blank, when summarizing all the scores of the 20 individual protocol stories, to represent the check marks on the individual *TAT* Scoring Blanks, the number of tally marks in the blank immediately before each characteristic or quality represents the number of times the subject has projected that particular characteristic or quality in his 20 protocol stories.

---

*Received in the Editorial Office on October 6, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] A section of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The Pennsylvania State College, January, 1952, an abstract of which was published in *The Journal of Psychology,* 1952, **34**, 27-30. Another section of this dissertation containing "*TAT* norms for males and females in a college population, and in a prison population," will be published immediately.

It should be noted that this scoring scheme was designed to indicate the tendency of a subject to manifest certain empathies and/or to project onto others the affective and personality qualities which are experienced by that subject himself. If the subject's tendencies are indicative of pathological conditions, then he may be called in for an inquiry based on his protocol stories and thus a more detailed evaluation of his pathological trends could be made.

B. Definition of Terms

1. *Press*

A press may be defined as the resultant of those forces emanating from the environment and acting upon a subject. It may be further defined as that effect which another person, thing, object, condition, or situation, either singly or in any combination thereof, exerts, may exert, or could exert upon a given subject in either a positive or negative direction. Again, a press may be defined as a gestalt of temporal stimuli which usually appears in the form of a threat of discomfort and/or harm, or in the form of a promise of gratification and/or a benefit to the subject concerned.

2. *Need*

A need may be defined as those drives or forces emanating from within the subject himself and acting upon him. It may be further defined as that force in the brain region which organizes perception, apperception, intellection, conation, and action in such a way as to transform them into a coördinated force which is directed towards the achievement of physiological, emotional, and/or psychological equilibrium by remedying or eliminating an existing and unsatisfying condition or situation.

3. *Thema*

A thema may be defined as that goal or objective towards which a press-need resultant force is directed for the purpose of attaining physiological, emotional, and/or psychological equilibrium by remedying or eliminating an existing and unsatisfying condition or situation.

4. *Empathy*

Empathy may be defined as the act of one individual entering into the feelings which are manifested or indicated by another individual or object.

a. Empathy towards objects may be considered as being esthetic in nature and as being the imaginal or mental projection of oneself into the elements of a work of art, or into a natural object.

b. Empathy towards individuals may be considered as a mental state in which one individual identifies himself with or feels himself as being in the same situation, conditions, or state of mind in which he observes another individual or group of individuals to be.

C. GENERAL INSTRUCTIONS

1. Only those empathies, presses, needs, and themas which are present at the time of the scene or event depicted in the *TAT* picture are to be scored. Those which are interpolated and are attributed to the predominant character at some date or time period subsequent to that depicted in the *TAT* picture are not to be scored. For example: One subject, a male prisoner, described the predominant character in the *TAT* picture 3-BM as being: "A prisoner who is now in the death cell of a prison after he had been convicted of murder and subsequently sentenced to death in the electric chair, and who is distressed and in mental anguish because of the terrifying experiences through which the court has decreed he must undergo." Then the subject interpolated a subsequent event in which he portrayed the predominant character as being a successful business man, happily married, and leading a normal life with his family and children, due to the fact that his death sentence had been commuted to life imprisonment, and then at a later date he had received a full pardon from a subsequent Governor of the State in which the prison was located because the predominant character had manifested good behavior and had been a model prisoner. It should be noted that the first projection of feelings into and the interpretation of the picture were confined to the situation and time period indicated in the *TAT* picture; however, the subsequent interpolated projection of feelings and interpretation were not confined to the situation and time period indicated in the *TAT* picture, and, furthermore, there was no indication of the fact that the predominant character in the first projection and interpretation was even considering or thinking of the feelings and situation interpolated in the subsequent projection and interpretation. Since those empathies, presses, needs, and themas, which were interpolated and attributed to the predominant character at a period of time subsequent to that indicated in the *TAT* picture and thus could not have been foreseen or experienced by the predominant character at the time indicated by the *TAT* picture, they are in the realm of pure fantasy and are not limited by the facts and laws of reality, and for that reason they should not be scored; consequently, only those projections and interpretations which are confined to the situation and time period indicated in the *TAT* picture and thereby limited by the facts and laws of reality should be scored. It should be noted that the subject is limited by the

realities suggested by the picture as long as he confines his interpretations and projections to the particular situation and time event indicated in the picture, and he thus reveals his own tendencies and personality traits as he manifests them in real life situations; however, when the subject does not confine his interpretations and projections to the situation and time event suggested to him by the picture, then his projections become mere daydreams and are no more reliable indices of his personality traits and tendencies than any other daydream. Subjects, when they do not limit themselves to the situation and time event suggested by the *TAT* picture, as illustrated above, tend to project into the interpolated daydream type of situation those social behavior patterns, personality traits, and concepts of moral behavior which they think that society or their associates expect them to possess and display.

2. The subject reveals his philosophy, ideals, standards, and concepts of social and ethical conduct or behavior by projecting them onto one of the characters in the *TAT* picture whom he selects to play the predominant rôle. The relationship of this predominant character to each of the subordinate characters, and/or to all of the subordinate characters as a group may be stated specifically or implied by the subject in his protocol of the *TAT* picture concerned.

3. A subject customarily indicates the unusual or the extreme educational phases of life; therefore, in checking the "Educational Life Phase" of the predominant character designated in the protocol, the "High School Educational Life Phase" should be checked unless otherwise indicated by the subject.

D. Definition of Scoring Categories

The categories used in this scoring blank are defined below. The limitations of these definitions should be observed conscientiously so that the evaluations which may be derived therewith will be reliable and comparable. It should be noted that the divisions and subdivisions which are used in this *TAT* Scoring Scheme are used in the discussion below.

I. PREDOMINANT CHARACTER

The rôles which may be assigned to, or assumed by, the predominant character in the *TAT* protocol story are self-explanatory and should be checked accordingly. In checking the items, "Relative" and "Other," a notation should be made in the right-hand margin to indicate the relationship, nature, or type of individual the predominant character may have or be.

II. SUBORDINATE CHARACTER

The rôles which may be assigned to, or assumed by, the subordinate character, or characters, in the *TAT* protocol story are self-explanatory and should be checked accordingly. In checking the items, "Relative" and "Other," a notation should be made in the right-hand margin to indicate the relationship, nature, or type of individual the subordinate character, or characters, may have or be.

III. THE RESULTANT EFFECT UPON OR THE REACTION OF THE PREDOMINANT CHARACTER WITH RESPECT TO THE PREVAILING PRESSES, OR ENVIRONMENTAL FORCES, WHICH ARE ACTING UPON, OR STIMULATING HIM

1. *Physiological Aspect Manifested, or Encountered, by the Predominant Character, His Family or His Associates*

a. Physical Well-Being: Normal physical health and physiological functioning.

b. Physical Disorder: Physical malfunctioning due to a trauma, disease, or a chemical.

c. Deprivation: Deprivations of a limb, organ, or a sense modality.

d. Addiction State: Addiction to alcohol and/or narcotic drugs.

e. Death: Termination of an individual's physical life due to any or all causes other than suicide or murder.

f. Suicide: Termination of an individual's own physical life due to his own intentional act or acts.

g. Murder: Termination of an individual's physical life due to the malicious and intentional act or acts of another individual or group of individuals.

2. *Emotional Aspect Manifested by Predominant Character*

a. Equilibrium: Normal emotional equilibrium and adjustment.

b. Elation: A heightened or intensified state of delight or satisfaction.

c. Dejection: A moderate state of being downcast, dispirited, or discouraged.

d. Depression: A state of despondency, hopelessness, or profound discouragement.

e. Phobias: Exaggerated, abnormal, and/or morbid fears.

f. Compulsion-Obsession: An emotional state characterized by obsessive ideas, doubts, and compulsive ceremonial behavior of a complicated sort.

3. *Mental Aspect Manifested by Predominant Character*

a. Rational: Mental behavior that is based on reason and on logical principles.

b. Incoherent: Mental behavior characterized by a lack of orderly, logical, and systematic connection of mental behavior units.

c. Ideas of Reference: Impressions that the actions or statements of others have some reference to oneself, either directly or implied.

d. Delirious: Mental behavior characterized by confusion and excitement.

e. Delusional: Mental behavior characterized by false beliefs or senses that do not yield to reason.

f. Hallucinatory: Mental behavior characterized by a false perception having no relation to reality and which is not accounted for by any external stimuli.

g. Fantasy: Mental behavior characterized by daydreaming, wishful thinking, and a disregard for reality which one desires to escape by revelling in imaginative possibilities.

4. *Economic Aspect Manifested by Predominant Character*

a. Means:

(1). *Abundance:* An economic state characterized by the possession of economic means above the amount that is normally considered adequate for the individual's accepted social status.

(2). *Adequacy:* An economic state characterized by the possession of economic means that are normally considered as being adequate for the individual's social status.

(3). *Want:* An economic state characterized by the possession of economic means that are below the amount normally considered as being adequate for the individual's social status.

b. Activities:

(1). *Vocational:* Those activities, either positive or negative, which are below the professional level, and in which a person may be engaged for the purpose of earning economic means.

(2). *Professional:* Those activities, for which superior intelligence and a prolonged period of training are required, and in which a person is engaged for the purpose of earning economic means.

(3). *Avocational:* Those activities in which a person is engaged merely for the delight, pleasure, and satisfaction which he may derive therefrom.

5. *Social Aspect, Either Positive or Negative, Manifested by the Predominant Character*

a. Parental Phase: The duties, obligations, and relationships of a parent.

b. Filial Phase: The duties, obligations, and relationships of a son or a daughter.

c. Fraternal Phase: The duties, obligations, and relationships of siblings.

d. Affinity Phase: The duties, obligations, and relationships of family relatives, not of the immediate family, and/or those acquired by marriage.

e. Amorous Phase: The accepted or customary obligations and relationships of a lover to the beloved member of the opposite sex, or affianced person.

f. Marital Phase: The duties, obligations, and relationships of a husband or a wife, each to the other.

g. Moral Phase: The accepted or customary obligations and relationships of an individual to his fellow members of the community in which he lives.

h. Ethical Phase: The accepted or customary duties, obligations, and relationships of a professional person to his clients or patients, as well as to the fellow members of his profession.

6. *Religious Aspect Manifested by Predominant Character*

a. Indifferent: Behavior characterized by a disinterested, neutral, and an unbiased attitude towards religion.

b. Devout: Behavior characterized by a solemn or reverent attitude towards religious exercises or observances.

c. Zealous: Behavior characterized by enthusiasm, devotion, and an ardent and active interest in religious activities, observances, and/or objectives.

d. Fanatic: Behavior characterized by excessive zeal and/or abnormal enthusiasm for religious exercises or observances.

e. Atheistic: Behavior characterized by a denial or a disbelief in the existence of a Supreme Being or God.

f. Penitent: Behavior characterized by adherence to the ideal that repentance and contriteness, and a resolution to make amends are necessary to atone for sins committed.

g. Forgiving: Behavior characterized by adherence to the ideal that one must be merciful, understanding, tolerant, placable and willing to pardon and to cease to feel resentment against others for their unfriendly acts in order that he may receive Divine forgiveness.

IV. ATTITUDES, FEELINGS, OR MOODS MANIFESTED BY THE PREDOMINANT CHARACTER

Do the attitudes, feelings, or moods manifested by the predominant character indicate that he is experiencing a need that is characterized by being

1-*a. Aggressive:* Progressive, forward, dominant, courageous, daring, offensive and/or hostile?

1-*b. Defensive:* Resisting an attack or aggression either real or imaginary; offering reasons in justification of his behavior, attitudes, or ideas; and/or retreating or withdrawing from an aggressor either real or imaginary, and seeking defense in flight?

2-*a. Altruistic:* Regardful of his fellowmen, and unselfish in his behavior?

2-*b. Selfish:* Unduly or supremely interested in his own personal interests and welfare at the expense of those of his friends and fellowmen?

3-*a. Ambitious:* Quite eager to achieve honor, superiority, power, authority and/or the attainment of some objective?

3-*b. Indifferent:* Neutral with respect to the value or desirability of attaining, or achieving a given objective?

4-*a*. *Assurance:* Self-confident, self-reliant, certain of his expected objectives; free from doubt as to the outcome of his objectives?

4-*b*. *Anxiety:* Manifesting a concern, dread, fear, foreboding, misgiving, worry, solicitude, uneasiness, or disquietude with respect to the outcome of some event, either real or imaginary?

5-*a*. *Cheerful:* Lively, gay, animated, lightsome, gleeful, jolly, vivacious, happy, and/or buoyant?

5-*b*. *Morose:* Gruff, severe, austere, gloomy, crabbed, surly, moody, sullen, and/or ill-humored?

6-*a*. *Compliant:* Obedient, submissive, complaisant, yielding, and attentive to, and/or respectful of legal authority and custom?

6-*b*. *Rebellious:* Disobedient, defiant, stubborn, headstrong, unyielding, unruly, and/or intractable towards legal authority and custom?

7-*a*. *Cooperative:* Willing to concur spontaneously with others in action, effort, and/or undertakings of a legitimate nature?

7-*b*. *Antagonistic:* Unwilling to concur spontaneously with others in action, effort, and/or undertakings of a legitimate nature; and tends to be rather hostile, repugnant, inimical, discordant, inconsistent, and adverse?

8-*a*. *Frank:* Free, outspoken, direct, straightforward, honest, unreserved, and sincere?

8-*b*. *Evasive:* Successful in avoiding, shunning, or escaping commitments of himself through the use of skill, contrivance, and/or dexterity?

9-*a*. *Friendly:* Amicable, kind, considerate, pleasing, amiable, peaceable, harmonious, and cordial?

9-*b*. *Hostile:* Resentful, vindictive, adverse, contrary, repugnant, inimical, and revengeful?

10-*a*. *Humble:* Modest, meek, unpretending, unassuming, and mild?

10-*b*. *Arrogant:* Proud, disdainful, domineering, bold, forward, impudent, and impertinent?

11-*a*. *Industrious:* Steadily and perseveringly active, zealous, painstaking, diligent, and assiduous?

11-*b*. *Indolent:* Habitually idle, lazy, inactive, indulging in ease, and avoiding labor and exertion?

12-*a*. *Loyal:* Characterized by fidelity, faithfulness, constancy, and devotion to a person, organization, cause, and/or principle?

12-*b*. *Unfaithful:* Heedless, indifferent, careless, and negligent with respect to a person, organization, cause, and/or principle?

13-*a*. *Optimistic:* Characterized by a disposition to take the most hopeful view?

13-*b*. *Pessimistic:* Characterized by a disposition to take the least hopeful view?

14-*a*. *Persevering:* Characterized by persistence, steadfastness, steadiness, constancy, and/or pertinacity?

14-*b*. *Vacillating:* Characterized by a wavering in conduct or purpose, irresolution, changeableness, and/or instability?

15-*a*. *Responsive:* Ready or inclined to answer, reply, rejoin, and/or to be amenable?

15-*b*. *Apathetic:* Void of feeling, not susceptible of deep emotion, indifferent, passionless, and/or unemotional?

16-*a*. *Self-Reliant:* Self-confident; trusting to his own abilities, powers, and judgments; and reliant upon himself?

16-*b*. *Dependent:* Relying on someone else for support or encouragement, and unable to sustain himself, or to perform anything, without the will, power, or assistance of another individual?

17-*a*. *Self-Respect:* Characterized by respect for himself, laudable self-esteem, and/or regard for his character?

17-*b*. *Self-Reproach:* Characterized by rebuking, censuring, or blaming himself for certain qualities, and/or for the lack of certain other qualities?

18-*a*. *Skeptical:* Characterized by a critical or incredulous attitude towards his inquiries, and by doubt or suspended judgment concerning data encountered?

18-*b*. *Credulous:* Ready to believe on slight evidence, or to be uncritical of evidence presented to him?

19-*a*. *Sociable:* Affable, accessible, companionable, communicative, friendly, and fond of mingling or talking with others?

19-*b*. *Seclusive:* Retiring, withdrawn from others, desirous of solitude, and privacy?

20-*a*. *Tolerant:* Forbearing, indulgent, enduring, permissive, broad-minded, and free from prejudice?

20-*b*. *Bigoted:* Prejudiced, narrowminded, obstinate, and blindly and unreasonably devoted or attached to an opinion, creed, party, practice, ritual, and/or system?

V. LIFE PHASE OF THE PREDOMINANT CHARACTER AT THE TIME INDICATED IN THE STIMULUS PICTURE

1. *Physical Life Phase*

a. *Childhood Phase:* That physical life phase or chronological age period which is comprised of the first 12 years of chronological age.

b. *Adolescent Phase:* That physical life phase or chronological age period which occurs between the 12th and the 21st chronological age birthdays.

c. *Adult Phase:* That physical life phase or chronological age period which begins with the 21st chronological age birthday and ends with the death of the individual concerned.

2. *Educational Life Phase (Which Was Attained, or in the Process of Being Attained at the Time Indicated in the Stimulus Picture)*

a. *Grammar School:* That educational phase of life which is comprised of the first eight grades in schools approved by the State Department of Public Instruction.

b. High School: That educational phase of life which is comprised of the grades 9, 10, 11, and 12 in secondary schools approved by the State Department of Public Instruction.

c. Trade School: That educational phase of life which is comprised of a brief period of time during which the individual acquires those techniques and skills required by a specific vocational trade or skilled labor, and for which graduation from an accredited high school may or may not be required.

d. Technical School: That educational phase of life which is comprised of a moderately extended period of training in a school, for entrance into which graduation from an accredited high school is a prerequisite, which trains an individual for technical positions, but which does not confer baccalaureate degrees.

e. College Level: That educational life phase which is comprised of at least four years in qualifying for the baccalaureate degree, and still further additional periods of time required for the master and for the doctoral degrees, as may be prescribed by law, and by the requirements for classification as an accredited degree granting college.

3. *Industrial Life Phase*

a. Unemployed: That industrial phase of life in which an individual is not gainfully employed.

b. Employed, Unskilled: That industrial phase of life in which an individual is gainfully employed in a vocation that requires no specified skill.

c. Employed, Skilled: That industrial phase of life in which an individual is gainfully employed in a vocation that requires the acquisition of certain specific skills and techniques as prerequisites, but which are below the supervisory and professional levels.

d. Employed, Supervisory: That industrial phase of life in which an individual is gainfully employed in a vocation that requires the ability to supervise the activities of fellow employees of an organization, or to supervise the activities of a home.

e. Employed, Professional: That industrial phase of life in which an individual is gainfully employed in professional work, the standards for which are established by law and/or professional organizations.

VI. QUALITY OF THE PROTOCOL STORY

1. *Structure: Evaluate the protocol story with respect to its composition and indicate if it is:*

a. Simple: Written with the complexity of structure that is used by a child, or by an adult with subnormal intelligence.

b. Ordinary: Written with the complexity of structure that is used by an adult with normal intelligence.

c. Complex: Written with the complexity of structure that is commonly used by an adult with superior intelligence.

2. *Plot: Evaluate the protocol story with respect to the complexity of its plan or plot, by means of which the interrelationships of the characters in the story are achieved and revealed, and indicate if the plot is:*

a. Simple: Constructed with the complexity commonly employed by children with normal intelligence, or by adults with subnormal intelligence.

b. Ordinary: Constructed with the complexity commonly employed by adults with normal intelligence.

c. Complex: Constructed with the complexity commonly employed by adults with superior intelligence, and/or advanced educational achievements.

3. *Aspect: Evaluate the protocol story with respect to the aspect, or point of view and orientation, which was manifested by the predominant character, and indicate if it was:*

a. Bizarre: That aspect or point of view which deals with or stresses the exotic, the extravagant, and/or sensational contrasts involved in an event or attributed to the characters and objects in that event, and which are not commonly accepted or used.

b. Realistic: That aspect or point of view manifested by the predominant character which deals with reality and the relationships between items of reality that are commonly accepted by the members of the community in which he lives.

c. Fantastic: That aspect or point of view which deals with or is concerned about events, relationships, qualities, objects, and/or creatures which do not exist in reality, but which exist only in fancy or imagination of the predominant character.

VII. CONTINGENCY OF THE EVENT EXPRESSED IN THE PROTOCOL STORY

1. *Certain*

Does the predominant character consider the occurrence of the event in the protocol story to be definite, a reality, or certain to happen?

2. *Probable*

Does the predominant character consider the occurrence of the event in the protocol story as being an event that is somewhat likely to happen, but not certain to occur?

3. *Improbable*

Does the predominant character consider the circumstances to be of such a nature, and so unusual, that the event is most unlikely to occur, and cannot be expected to happen?

VIII. REACTION OF PREDOMINANT CHARACTER TO THE OUTCOME OF THE EVENT

(It should be noted that only the time event indicated in or suggested by the *TAT* stimulus picture is to be considered.)

1. *Accepted*

Has the predominant character accepted the present event that has occurred as being one that he desired or considered favorable to his welfare and objectives? If the predominant character did not consider the event as being one that he desired or considered favorable to his welfare, has he accepted the event as being the best possible one under the circumstances and has he adjusted himself to the situation?

2. *Endured*

Did the predominant character consider the present event as being one that was not favorable to his welfare and objectives, but one that must be endured by him because the factors involved in the event were beyond his control?

3. *Rejected*

Did the predominant character reject the present event because he considered it as being one that was unfavorable to his welfare and to the achievement of his objectives, and attempt to find a solution to the problem either through his own efforts or through the efforts of others?

IX. PREVAILING THEMAS, GOALS OR OBJECTIVES WHICH WERE MANIFESTED BY THE PREDOMINANT CHARACTER, AND WHICH WERE CONCERNED WITH HIS ACQUISITION AND/OR MAINTENANCE OF

1. *Economic Goals or Objectives*

a. Monetary assets, e.g., cash, stocks and bonds.

b. Material possessions, e.g., real estate, personal property, livestock, and/or machinery.

c. Indifference to all economic objectives.

2. *Prestige Goals or Objectives*

a. Political prestige or position.

b. Social prestige and/or acceptance in his or her immediate community.

c. Economic prestige or position, e.g., economic prestige derived or attained by becoming an official in a corporation or financial institution where he can influence or regulate the amount of salaries paid, or the amount and quantity of loans issued.

d. Scholastic prestige, e.g., prestige derived by being an authority in a particular subject or field, such as: engineering, religion, science, ceramics, etc.

e. Indifference to all prestige goals or objectives.

3. *Welfare Goals or Objectives Which Are Concerned with*

a. Himself, personally.

b. His own immediate family or relatives.

c. His friends.

d. The community in which he lives.

4. *Intellectual Attainments*

a. The basic intellectual skills and information provided by the public schools of the high school or lower.

b. The artistic skills and proficiency required in the work of a portrait and/or landscape painter, a commercial artist, a commercial photographer, and similar occupations.

c. The literary skill and proficiency required in journalistic work, and in writing fiction and poetry.

d. The skill and proficiency in theoretical subject matter required for entrance into the scientific pursuits, and/or research, e.g., electronics, chemistry, physics, etc.

e. The skill and proficiency in theoretical subject matter, as well as in the professional techniques, which are prescribed by law and professional organizations for entrance into the recognized professions, e.g., an educator, lawyer, physician, registered nurse, etc.

f. Indifference as to the attainment of a higher intellectual status.

g. The skill and proficiency in theoretical subject matter, as well as economic and business administration techniques, which are required for employment in the economic and business administration vocations.

5. *Vocational Pursuits*

a. Unskilled labor activities.

b. Skilled labor activities.

c. Technical activities which may or may not require a college degree, e.g., an accountant, a commercial artist, a commercial photographer, etc.

d. Professional activities, the prerequisites for entrance into which are prescribed by law and the respective professional organizations.

e. Indifference to employment in any gainful occupation of a legitimate nature.

f. Managerial activities required in a business organization, or in a home.

6. *Avocational Pursuits*

a. Field sports, e.g., baseball, football, fishing, hunting, etc.

b. Artistic activities, e.g., artistic painting and/or drawing, engraving, photography, etc.

c. Craftsmanship, e.g., leathercraft, woodcraft, mechanical repairing, etc.

d. Fads, e.g., nudist colony activities, physical culture groups, etc.

e. Entertainment, e.g., various recreational, card and intellectual games; amateur acting, motion pictures, amateur magician, etc.

f. Music, e.g., vocal and/or instrumental.

g. None, i.e., no avocational activity was indicated.

7. *Family Goals or Objectives*

a. Acquisition of a:

(1). Spouse.

(2). Child.

b. Discharge of His Obligations as a:

(1). Parent.

(2). Sibling.

(3). Spouse.

(4). Child.

(5). Relative other than those of his own immediate family, which also includes the in-laws acquired by marriage.

(6). Indifferent to the duties and obligations to the members of his family, his relatives, and his in-laws as required by custom and/or law.

8. *Sensual Gratification, i.e., the Immediate Gratification of the Natural or Acquired Demands for*

a. Bodily sustenance, e.g., food and water.

b. Intoxicants, e.g., wines, whiskey, etc.

c. Narcotics, e.g., heroin, morphine, opium, etc,

d. Sex, e.g., any sexual activity, normal or perverted, legitimate or illegitimate.

e. Tobacco, e.g., the use of tobacco in any form.

f. Sublimation, e.g., sublimation of a strong sensual impulse by active participation in official activities, religious activities, etc.

g. None indicated.

9. *Religious Goals or Objectives*

a. Piety, e.g., manifesting a reverence for God, the Supreme Being, and discharging his religious obligations as prescribed by religious creed.

b. Penitence, e.g., manifesting repentance and a contriteness and a feeling of sorrow for his sins and offenses.

c. Forgiveness, e.g., manifesting a desire for God to forgive and pardon him for the sins he has committed, or believes he has committed.

d. Mission, e.g., manifesting an attempt to accomplish some religious duty or objective for which he considers himself obligated.

e. Indifference to all religious matters.

10. *Fantasy, Those Goals or Objectives Characterized by*

a. Delusions, false beliefs which do not yield to reason.

b. Hallucinations, false perceptions not accounted for by any external stimuli.

c. Daydreaming, wishful thinking, a disregard for reality which one desires to escape by revelling in imaginative possibilities.

d. Constructive imagination, logical imagination based on reality and directed constructively.

e. None stated or indicated by the predominant character.

X. DIRECTION OF VECTORS

1. *Positive*

Those themas, goals or objectives which are expressed by the predominant character and which are concerned with the acquisition or achievement of objectives that are permitted or authorized by law, and which have been accepted by the community in which the predominant character lives.

2. *Negative*

Those themas, goals or objectives which are expressed by the predominant character and which are concerned with the acquisition or achievement of objectives that are prohibited by law, and which have been disapproved or rejected by the community in which the predominant character lives.

180 South River Street
Wilkes-Barre, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 197-200.

TAT SCORING BLANK*[1]

VA Regional Office, Wilkes-Barre, Pennsylvania

FRANKLYN D. FRY

This *TAT* Scoring Blank was developed by the author to facilitate the scoring of the *TAT* protocol stories according to the scoring scheme that the author developed for the evaluation of the *TAT* protocol stories.[2]

It should be noted that this *TAT* Scoring Blank may be used to score just one protocol story, or it may be used to summarize the scores of all of the 20 individual protocol stories. When scoring an individual protocol story, the number of the *TAT* picture, about which the protocol story has been written, should be inserted in the blank after the expression, "*TAT* Picture," but when it is used to summarize the scores of all of the 20 individual protocol stories, the term, "Summary," should be written in the blank after the expression, "*TAT* Picture."

In evaluating an individual protocol story, the predominating characteristic or quality should be checked in the blank that immediately precedes that characteristic or quality. Under the heading of I only one characterization should be checked, since there is only one predominant character. Under the heading of II, since there may be no subordinate character at all or there may be one or more subordinate characters, the characterizations should be checked as indicated in the protocol story. Under the heading of III only one characteristic or quality should be checked in each of the sub-divisions, 1, 2, 3, 4-a, 4-b, 5, and 6, since it is the predominant characteristic or quality that should be checked. Under the heading of IV the predominant characteristic or quality in each of the 20 dichotomies should be checked. Under the headings of V, VI, and IX one appropriate characteristic or quality in each of the numbered sub-divisions should be checked. Under each of the headings VII, VIII, and X only one appropriate characteristic or quality should be checked.

When this *TAT* Scoring Blank is used to summarize the evaluations of the 20 individual *TAT* protocol stories, tally marks should be placed in the blank immediately preceding the appropriate characteristic or quality that

---

*Received in the Editorial Office on October 6, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] A section of a dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy, The Pennsylvania State College, January, 1952, an abstract of which was published in *The Journal of Psychology,* 1952, **34**, 27-30.

[2] Published in *The Journal of Psychology,* 1953, **35**, 181-195.

had been checked on the *TAT* Scoring Blanks used for the evaluation of each of the individual protocol stories. Then, after all of the individual evaluations have been tallied, the total tally marks should be as follows:

1. Under each of the headings of I, VII, VIII, and X there should be a total of 20 tally marks.
2. Under the heading of III there should be a total of 20 tally marks in each of the subdivisions, 1, 2, 3, 4-a, 4-b, 5, and 6.
3. Under the heading of II the total tally marks may be any number.
4. Under each of the headings of V and VI there should be a total of 20 tally marks in each of the subdivisions, 1, 2, and 3.
5. Under the heading of IV there should be a total of 20 tally marks for each of the 20 dichotomies.
6. Under the heading of IX there should be a total of 20 tally marks for each of the numbered sub-divisions.

TABLE 1
TAT SCORING BLANK

I. PREDOMINANT CHARACTER

1.—Father.
2.—Mother.
3.—Brother.
4.—Sister.
5.—Husband.
6.—Wife.
7.—Son.
8.—Daughter.
9.—Relative.
10.—Self.
11.—Other.

II. SUBORDINATE CHARACTER

1.—Father.
2.—Mother.
3.—Brother.
4.—Sister.
5.—Husband.
6.—Wife.
7.—Son.
8.—Daughter.
9.—Lover.
10.—Admirer.
11.—Friend.
12.—Enemy.
13.—Relative.
14.—Self.
15.—Other.

III. THE RESULTANT EFFECT UPON OR THE REACTION OF THE PREDOMINANT CHARACTER WITH RESPECT TO THE PREVAILING PRESSES, OR ENVIRONMENTAL FORCES, WHICH ARE ACTING UPON, OR STIMULATING HIM

1. *Physiological Aspect:*
   a.—Physical Well-Being.
   b.—Physical Disorder.
   c.—Deprivation.
   d.—Addiction State.
   e.—Death.
   f.—Suicide.
   g.—Murder.

2. *Emotional Aspect:*
   a.—Equilibrium.
   b.—Elation.
   c.—Dejection.
   d.—Depression.
   e.—Phobias.
   f.—Compulsion-Obsession.

3. *Mental Aspect:*
   a.—Rational.
   b.—Incoherent.
   c.—Ideas of Reference.
   d.—Delirious.
   e.—Delusional.
   f.—Hallucinatory.
   g.—Fantasy.

4. *Economic Aspect:*
   a. *Means:*
      (1)—Abundance.
      (2)—Adequacy.
      (3)—Want.
   b. *Activities:*
      (1)—Vocational.
      (2)—Professional.
      (3)—Avocational.

5. *Social Aspect:*
   a.—Parental Phase.
   b.—Filial Phase.
   c.—Fraternal Phase.
   d.—Affinity Phase.
   e.—Amorous Phase.
   f.—Marital Phase.
   g.—Moral Phase.
   h.—Ethical Phase.

6. *Religious Aspect:*
   a.—Indifferent.
   b.—Devout.
   c.—Zealous.
   d.—Fanatic.
   e.—Atheistic.
   f.—Penitent.
   g.—Forgiving.

TABLE 1 (*continued*)

IV. ATTITUDES, FEELINGS, OR MOODS MANIFESTED BY THE PREDOMINANT CHARACTER

1.	a.—Aggressive,	b.—Defensive.	11.	a.—Industrious,	b.—Indolent.
2.	a.—Altruistic,	b.—Selfish.	12.	a.—Loyal,	b.—Unfaithful.
3.	a.—Ambitious,	b.—Indifferent.	13.	a.—Optimistic,	b.—Pessimistic.
4.	a.—Assurance,	b.—Anxiety.	14.	a.—Persevering,	b.—Vacillating.
5.	a.—Cheerful,	b.—Morose.	15.	a.—Responsive,	b.—Apathetic.
6.	a.—Compliant,	b.—Rebellious.	16.	a.—Self-Reliant,	b.—Dependent.
7.	a.—Coöperative,	b.—Antagonistic.	17.	a.—Self-Respect,	b.—Self-Reproach.
8.	a.—Frank,	b.—Evasive.	18.	a.—Skeptical,	b.—Credulous.
9.	a.—Friendly,	b.—Hostile.	19.	a.—Sociable,	b.—Seclusive.
10.	a.—Humble,	b.—Arrogant.	20.	a.—Tolerant,	b.—Bigoted.

V. LIFE PHASE OF THE PREDOMINANT CHARACTER

1. *Physical:*	2. *Educational:*	3. *Industrial:*
a.—Childhood.	a.—Grammar School.	a.—Unemployed.
b.—Adolescent.	b.—High School.	b.—Employed, Unskilled.
c.—Adult.	c.—Trade School.	c.—Employed, Skilled.
	d.—Technical School.	d.—Employed, Supervisory.
	e.—College Level.	e.—Employed, Professional.

VI. QUALITY OF STORY

1. *Structure:*	a.—Simple,	b.—Ordinary,	c.—Complex.
2. *Plot:*	a.—Simple,	b.—Ordinary,	c.—Complex.
3. *Aspect:*	a.—Bizarre,	b.—Realistic,	c.—Fantastic.

VII. CONTINGENCY OF EVENT

1.—Certain. 2.—Probable. 3.—Improbable.

VIII. REACTION OF PREDOMINANT CHARACTER TO THE OUTCOME OF THE EVENT

1.—Accepted. 2.—Endured. 3.—Rejected.

IX. PREVAILING THEMAS, GOALS OR OBJECTIVES MANIFESTED BY THE PREDOMINANT CHARACTER, WHICH ARE CONCERNED WITH HIS ACQUISITION AND/OR MAINTENANCE OF

1. *Economic Goals or Objectives:*
   a.—Monetary Assets b.—Material Possessions. c.—Indifference to All Economic Objectives.
2. *Prestige Goals or Objectives:*
   a.—Political. b.—Social. c.—Economic. d.—Scholastic.
   e.—Indifference to All Prestige Goals or Objectives.
3. *Welfare Goals or Objectives:*
   a.—Personal. b.—Family. c.—Friends. d.—Community.
4. *Intellectual Attainments:*
   a.—Basic Intellectual Skills and Information. b.—Artistic. c.—Literary.
   d.—Scientific. e.—Professional. f.—Indifference. g.—Economic.
5. *Vocational Pursuits:*
   a.—Unskilled. b.—Skilled. c.—Technical. d.—Professional.
   e.—Indifference to All Legitimate Gainful Employment.
   f.—Managerial, Business or Home.

TABLE 1 (*continued*)

---

6. *Avocational Pursuits:*
   a.—Field Sports. b.—Artistic. c.—Craftsmanship. d.—Fads.
   e.—Entertainment. f.—Music. g.—None Indicated.
7. *Family Goals or Objectives:*
   a. *Acquisition of a:* (1)—Spouse. (2)—Child.
   b. *Discharge of His Obligations as a:* (1)—Parent. (2)—Sibling. (3)—Spouse. (4)—Child. (5)—Relative. (6)—Indifference to the Discharge of His Duties as a Family Member.
8. *Sensual Gratification:*
   a.—Bodily Sustenance. b.—Intoxicants. c.—Narcotics. d.—Sex. e.—Tobacco. f.—Sublimation. g.—None Indicated.
9. *Religious Goals or Objectives:*
   a.—Piety. b.—Penitence. c.—Forgiveness. d.—Mission.
   e.—Indifference to all religious activities.
10. *Fantasy, Goals or Objectives Characterized by:*
    a.—Delusions b.—Hallucinations. c.—Daydreaming. d.—Constructive Imagination. e.—None Stated.

X. DIRECTION OF VECTORS

1. *Positive:* —Acquisition of Legitimate Objectives.
2. *Negative:* —Acquisition of Illegitimate Objectives.

---

180 South River Street
Wilkes-Barre, Pennsylvania

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 201-226.

PROBLEM-ACTION RESPONSES AND EMOTIONS IN THEMATIC APPERCEPTION TEST STORIES RECOUNTED BY ALCOHOLIC PATIENTS*[1]

Payne Whitney Psychiatric Clinic and Hunter College

C. A. KNEHR, A. VICKERY, AND M. GUY[2]

A. THE PROBLEM

In connection with an extensive study of alcoholism being carried out in the Payne Whitney Psychiatric Clinic, we have attempted to determine by means of the Thematic Apperception Test (*TAT*) whether personality dynamics of alcoholic patients differ from those of other patients hospitalized in this Clinic.

The interest in this problem is two-fold. First, we have been interested in the general problem of whether the *TAT* provides the clinician with dynamic material which he might not otherwise obtain by usual clinical means. Further, assuming that such is the case, would the amount of data thus provided be such that the lengthy expenditure of time in using this test procedure be commensurate with the value of the dynamic leads obtained? It is obvious that this is in part an attempt to evaluate the usefulness of the *TAT* with hospitalized psychiatric patients. In recent papers (1, 2, 3, 5, 14) one observes a growing sense of discouragement with the results obtained with this test.

The second and primary area of interest lies in the determination of possible distinguishing personality characteristics of alcoholics which might be studied in relation to the etiology of chronic alcoholism. In any consideration of dependence on alcohol one ultimately will need to know something about the factors which lead to excessive use of alcohol in one individual and not in another. It seemed that an examination of *TAT* protocols for alcoholics as a group might reveal some tendencies which differ from those of nonalcoholic patients with respect to stated conflicts, situations, and the way in

---

*Received in the Editorial Office on October 13, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Study from The New York Hospital and the Department of Psychiatry, Cornell University Medical College, New York, and Hunter College of the City of New York.

[2] The authors are grateful to Marion R. Knehr for the preparation of the charts used in this paper.

which these are met. Further, would the differences in emotional responses seen clinically be reflected in the *TAT* stories?

The use of the *TAT* in a problem of this sort entails certain difficulties which can be stated in advance. To anyone experienced with the *TAT*, it is well known that it is a highly subjective technique from the point of view of interpretation, and will vary on this score with the theoretical orientation of the examiner. Many factors go into an interpretation of 20 stories in order to indicate the dynamics of the patient. It seems that one must adopt a shifting set of assumptions in extracting the dynamic content of stories for a given patient. One is continually forced to make choices of theoretical orientation since it may be difficult to determine whether the story represents analogous dynamics in the patient or perhaps represents a symbolic response to the story situation such that mechanisms of wish-fulfillment, sublimation, and the like are responsible for the kind of story told.

In routine use of the *TAT* we have employed a more behavioral framework for analysis. We have not assumed that we were penetrating to the level of unconscious repressed tendencies, nor have we interpreted fantasy material symbolically in terms of defensive and substitutive mechanisms of psychoanalytic theory (11). Dynamics are derived more directly from stated or implied thoughts and behaviors of the individuals who people the stories. Actually the symbolic approach has not been ruled out entirely, for whenever useful information was obtained by this means, the physician in charge of the patient was advised of the content and possible meaning. For the purpose of the present study, however, it seemed that a more uniform behavioral procedure was needed in order to arrive at any kind of generalization.

We are well aware that many individuals object to an attempt to make methods of personality study or techniques such as the *TAT* more objective or quantitative on the ground that in dealing with personality by this method, one tends to lose the detail and subtleties of the personality. It is, however, a characteristic of the scientific approach that there is always the sacrifice of a certain amount of detail and specificity in the interest of generalization. The scientific procedure of cutting through a mass of data to obtain the abstractions has been extremely important and productive in other fields. This may be all the more necessary in the field of personality study where elusive characteristics are sought in a maze of complexities.

The present study is a partially quantitative comparison between groups in which the individual story responses to the *TAT* are the basic data. This may be the first published study of protocols used in this manner, but it by no means exhausts the possibilities for ordering the mass of data presented in the test protocols of our patient groups. If some personality characteristics common to alcoholics could be shown by this procedure, the psychologist and psychiatrist would move one step nearer to an understanding of this psychopathological entity.

B. PROCEDURE

The subjects of this study were 84 inpatients of the Payne Whitney Psychiatric Clinic. Of this number, only 78 will appear in the intergroup comparisons. Six patients were omitted because three gave incomplete protocols, while three others received the children's form of the test.

Of these 78 patients, 33 (21 males and 12 females) showed alcoholism as a major element in the syndrome, while 45 (23 males and 22 females) were nonalcoholic patients with assorted difficulties. Throughout the remainder of this paper the two groups will be referred to as alcoholic and assorted patients respectively.

The procedure for the administration of the *TAT* is that generally used and recommended by Murray (10), Rotter (13) and others, with full instructions to indicate what led up to the situation, the present situation in the picture, the thoughts and feelings of the characters, and the outcome. The full series of 20 pictures was given to each patient, and the test was given in two sessions except in a few instances where three or more sessions were required. The test was administered during the first month of hospitalization and was generally given to the alcoholic group 21 days after admission.

In order to compare the dynamics obtained by the use of the *TAT* with those revealed in the course of routine clinical procedures in the hospital, individual protocols were typed and an analysis sheet appended. This latter included a paragraph describing the behavior of the patient during the test sessions and a listing of the dynamics, showing the frequency of the items by noting the cards in which the items appeared. Ordinarily, there was no interpretation of the test data in the usual sense. However, if an interpretation beyond the listed dynamics was deemed useful to the physician, it was included.

Whenever an analysis of the *TAT* protocol was submitted to the patient's physician, a request was made for the physician to note on the analysis sheet any listed dynamics which were not known to him at the time of testing.

This information, returned by the physician, provides the basis for the comparisons shown in the first portion of the results.

In order to cut through the mass of data provided by nearly 1,600 stories told for the *TAT* pictures and to arrive at a basis for group comparisons, two check sheets were prepared and may be seen in Figures 1 and 2. A number of classification schemes might have been chosen, but those illustrated appeared a priori to offer the best possibilities for comparing alcoholic with nonalcoholic patients. Once the decision had been made to classify each story with respect to the kind of problem, pressure, or situation and in terms of the actions taken by the story characters to deal with these problems, the selection of classifications under these headings became the next task.

The items appearing on the problem-action check sheet (Figure 1) were selected after reading and discussing the problems and action responses in several hundred stories chosen at random. When these items were compiled, we seemed to have a manageable two-way classification in which every story could be located, with sufficient divisions to insure adequate differentiation of stories. An inspection of Figure 1 will reveal that the 11 problem headings can be combined into five general classifications. The first main classification involves interpersonal relations and includes conflicts between persons of the opposite sex, the same sex, or with two other persons in a sex triangle. The next general group deals with problems of ascendance-sub-

	Heterosexual Conflicts	Homosexual Conflicts	Sex Triangle	Conflict with Authority	Use Made of Power	Situational Hope	Situational Fate	Ethical Standards	Personal Adequacy	Alternative Courses of Action	None Indicated
Aggressive Attack											
Retribution or Revenge											
Constructive Aggression											
Rational Approach											
Sensuous Gratification											
Compromise											
Seeks Help											
Passive Hostility											
Anxious Suspension of Activity											
Acceptance of Situation											
Irrational, Ill-directed Activity											
Submission											
Avoidance											
Resignation											
Active Withdrawal											
Suicide											
No Action											
Alternative Actions (Unresolved)											

FIGURE 1
PROBLEM-ACTION CHECK SHEET

mission (also, but not always, interpersonal). "Conflict with authority" implies an individual in the subordinate rôle, and "use made of power" refers to the character in the dominant rôle. The third general classification includes problems involving external stresses not occasioned by other people in an immediately personal way. The subheading "situational hope" implies a situation alterable by human efforts, while one designated as "situational fate" denotes situations beyond the control of characters in the stories. Conflicts involving "ethical standards," "personal adequacy," and "alternative courses of action" are intrapersonal problems covering a range of intensities in emotional involvement of the character. The final classification of "none indicated" includes patients' responses in which no plot or conflict appeared.

Referring again to Figure 1, 18 classifications of actions taken by story characters are shown in the left hand column. It will be seen that this represents a behavior spectrum ranging from the strongly positive, "moving into the situation," at the top, through a relatively neutral series of responses, to actively negative behaviors of withdrawal, escape, and suicide. "No action" is taken to mean the avoidance of the problem by the patient telling the story or description of a problem with no stated solution. This classification has been grouped with the negative behaviors.

When this check sheet was in preparation, much attention was given to insure the mutual understanding by the authors of the meanings of the various classifications and the kinds of problems and behaviors which would be included under them. In the actual procedure, one of us (A. V.) analyzed the large share of the stories. There was, however, random checking by the other authors throughout the analyses. Wherever there were stories which were difficult to classify, location on the check sheet was reached on the basis of discussion and mutual agreement. It is our opinion that this procedure has led to a consistency of interpretation that might otherwise have been lost. It is interesting to note that the stories most difficult to classify were from the protocols of patients with schizophrenic features in their pathology. Their responses had an evanescent, roundabout quality, and were relatively noncommittal regarding alternatives. This difficulty increased where the content became bizarre and fantastic.

While the individual analysis of dynamics was prepared and presented to the physician at the time each *TAT* was given, all stories were later reanalyzed at one time using the check sheet described above. The complete review of problem-action responses in all stories was made as follows. All 20 stories for a given patient were recorded on a single check sheet. For example, on Card 1, a story about a child submitting to his mother's wishes that he practice the violin when he would rather be playing baseball with his friends, would be noted on the check sheet under the column headed "conflict with authority" and in the row entitled "submission." It would be located by number (1) in this cell. In like manner, the remaining 19 stories of the individual protocol were analyzed and recorded. One could then see from the scatter of entries whether there was some general tendency in the patient to tell stories involving particular kinds of problems and whether the patient showed some characteristic pattern in meeting these problems. This procedure was followed for each of the 78 patients with due care that each story be located in accordance with the chief problem and the major type of action ascribed to the character involved.

These individual check sheets provided the basis for further analysis either by card or by patient group, depending on how the data were combined. For example, the problem action distribution of all alcoholic males to Card 4 could be obtained by extracting all Card 4 locations for all individuals of the group and locating them on a separate check sheet. A similar procedure to extract Card 4 locations for all female alcoholics would provide a second problem-action pattern, so that now the type of story and character responses to this card could be compared for the two sexes.

The results below represent varying analyses of this kind and were carried out at first by keeping the four sets of data separate. These groups were alcoholic males, alcoholic females, assorted males, and assorted females. Later these data were combined as to problem and response frequencies to all cards, maintaining the identity of the groups, with subsequent merging of the data on males and females of each of the two patient groups for certain comparisons.

The second check sheet (Figure 2) was designed primarily to note the presence or absence and the degree of anxiety, tension, and resentment attributed to the story characters by the patients in the study. The choice of emotional states was made to correspond to those being evaluated in the patients in another and independent study (7). A two-dimensional classification was made which would show not only the emotional state of the story characters, but also the situations or persons to which the emotions were referred. The categories are similar to the situational classifications in the problem-action check sheet discussed above.

In order to insure consistency in the evaluation of the emotional responses, we recorded only those responses which were actually ascribed to the story characters. We had to distinguish emotions that *could* have been experienced in a given situation from those which *were* indicated to be the response

	ANXIETY			TENSION			RESENTMENT			
Emotional Displacement Related to	Mild	Mod.	Mark.	Mild	Mod.	Mark.	Mild	Mod.	Mark.	None
Self										
Same sex-equals										
Other sex-equals										
Underlings										
Authority										
Society										
Situation										
Not indicated										

FIGURE 2
EMOTIONAL-DISPLACEMENT CHECK SHEET

made by the story characters. These should not be confused with the emotional state of the patient during examination.

As in the previous work sheet, the authors discussed the stories for a number of patients so that there was mutual understanding of the procedure, particularly with regard to the presence or absence of emotional responses in the story character, and agreement as to the degree of emotion represented. The situations to which the emotional responses were referred presented the least of the problems encountered.

In order to make intergroup comparisons, the data on emotional responses recorded on the individual check sheets were combined and analyzed in the same manner described for the problem-action data.

Since there are possible sex differences in responses to the pictures of the *TAT,* it is necessary that comparisons be made on stimulus cards common to all groups, both male and female. These common cards are 1, 2, 4, 5, 10, 11, 13, 14, 15, 16, 19, and 20. The analyses reported in the results include only the data on these 12 pictures and represent a common base for evaluation of any differential results.

C. RESULTS

The clinical usefulness of the *TAT* is perhaps best gauged by a comparison of the important personality characteristics obtained by the use of the test and the personality dynamics revealed in the course of therapeutic sessions during the early weeks of treatment. During this time the patients had been under relatively intensive study by the physicians, about three hours per week being spent in therapeutic interviews. In addition to this, a psychiatric examination of the patient was made by the physician, and he gathered information from sources other than the patient. In the individual analyses for 84 patients, the physicians noted new dynamic material unknown to them at the time of testing in only 14 patients, or about one patient in six. The true picture is not given by the figures, however, since most of the new dynamics reported as a result of testing were considered by the physicians to be relatively unimportant from the point of view of either understanding the patients' pathologies or formulating therapeutic procedures. Insofar as we are able to determine, only one of these patients revealed material on the *TAT* which affected the physician's insight into the case and his approach to it. The new material on the other 13 patients usually involved one or two characteristics, some of which were considered new merely in the degree to which they were emphasized. Physicians' statements indicated, by and large, that the *TAT* analyses confirmed rather than supplemented dynamics revealed early in the course of therapy.

An obvious use of the projective test in the clinical situation is in eliciting personality characteristics kept hidden by the patient in therapeutic interviews. Our experience with examinations, requested with just this purpose in mind, has been uniformly disappointing. Usually the patient who has his defenses firmly established for the clinical interview yields very little in the course of testing with Murray's pictures.

In the second portion of this study, namely the search for personality dynamics which would differentiate alcoholics from other hospitalized patients, we have made use of the story classifications described above.

It seemed possible that there would be sex differences in the kinds of problems suggested by the common cards, and the actions through which the story characters were moved. Card by card comparisons between male and female alcoholics, and again between assorted males and females, showed a surprisingly similar distribution of problem frequencies, which varied, however, from one card to the next. The card showing the greatest lack of similarity was Card 13 where the most frequent problem indicated by the males was one of ethical standards, whereas the females tended to see in this card a situation of fate or sexual conflict.

In Table 1, where we have combined the data for the 12 *TAT* cards common to the protocols for males and females alike, it is shown that the problem responses fail to reveal any significant sex difference. The consistency of the two sexes in problem-frequency patterns for the eleven classifications of the separate cards carries over to the combined data where

TABLE 1

THE SIMILARITY IN FREQUENCY OF PROBLEMS INDICATED BY ALCOHOLIC AND ASSORTED MALES AND FEMALES TO THE CARDS COMMON TO BOTH SEXES
(Frequencies are expressed as per cent.)

	Assorted patients		Alcoholic patients	
Problems	Males per cent	Females per cent	Males per cent	Females per cent
Interpersonal				
Heterosexual Conflicts	9.4	10.6	13.1	12.5
Homosexual Conflicts	1.5	1.1	.8	.7
Sex Triangle	4.7	3.4	4.4	1.4
Power				
Conflict with Authority	11.2	8.3	8.7	8.3
Use Made of Power	4.0	4.9	4.0	4.2
Situational				
Situational Hope	25.4	23.5	22.6	27.8
Situational Fate	6.9	5.7	6.3	9.7
Intrapersonal				
Ethical Standards	6.2	3.4	5.1	3.5
Personal Adequacy	3.6	6.1	6.0	9.7
Alternative Courses of Action	2.5	4.2	4.8	2.8
None Indicated	24.6	28.8	24.2	19.4
Chi Square*	2.13		3.55	
P	> .70		> .40	

*Chi squares are calculated by combining data into the five major problem headings above because of low frequencies in some cells.

the frequencies are expressed as per cent of the total. While the two male-female comparisons within the patient groups are highly similar, it is also evident that there is a similarity between assorted and alcoholic patients, although somewhat more variable. The chi squares indicate that there is no significant difference between the sexes of either group from the point of view of types of problems elicited by the 12 cards. In our opinion, these data are an adequate basis for combining the responses of alcoholic males and females, and assorted males and females for comparison of patient groups as a whole.

The differing distributions of plot frequencies from one card to another provide us with one of the more interesting sets of observations of this study which refer more to the *TAT* itself than to comparisons of results obtained with it.

An examination of the problem-action distributions for the separate cards suggests that each card tends to produce a pattern of its own with respect to the incidence of the kinds of problems the story characters face. The strength of the cards in determining the nature of the problems varies considerably, but gives the impression of having a rather marked controlling influence. Any intergroup comparisons must, therefore, be interpreted in the light of this controlling influence.

An examination of the same data shows a different situation with respect to what the story characters are caused to do in the face of the situations presented. Here the cards have a greatly decreased controlling value with story character responses varying rather widely from patient to patient. Thus, the *TAT* pictures appear to be relatively structured as to the nature of the plots suggested, and much less structured in terms of the actions ascribed to the characters.

Supporting data for these assertions are shown in Tables 2 and 3. The distribution of plot frequencies in Table 2 and the response frequencies in Table 3 are for all 78 patients. An examination of the first of these distributions shows that the cards differ among themselves in structuring. Card 4, for example, is highly structured in that 87 per cent of all patients told a story about interpersonal sexual problems. It is obvious that sexual content for this card has no discriminative value. Note also the distribution for Card 13. Here, nearly three-quarters of the plots involved ethical standards, heterosexual conflicts, or situations of a fateful nature. And notice the failure of patients to provide a story for Card 19. About 52 per cent of all stories for this card were without a plot or problem and tended to be merely descriptive in nature. Card 5, on the other hand, shows less concentration in a given type of problem and may therefore be considered to be less structured.

The distributions in Table 3 show the more variable and scattered behavior responses appearing in the stories for the common cards. It will be noted, for example, that Card 4, which elicited mostly interpersonal sexual problems, yields a less marked piling up of responses with greater scatter among the 18 categories of response. Card 13 shows perhaps more clearly the pronounced variability in response along with somewhat restricted theme types. A study of Card 19, for which the majority of the stories contained no plot, reveals that the greatest per cent of responses fall in the "no action" category, although beyond that there is an even distribution among the remaining types of actions. Just as Card 5 provided the least structured problem pattern, this card also allows greater freedom in the choice of behavior response. On the other hand, some of the cards evoke a relatively large incidence of responses in a narrowed field of activities. Examples are Cards 2 and 11. The latter shows a rather interesting effect. Action by

TABLE 2

VARIATIONS IN PROBLEMS OR THEMAS INDUCED BY DIFFERENT *TAT* CARDS SHOWN BY THE DISTRIBUTION OF FREQUENCIES EXPRESSED AS PER CENT*

(78 patients)

Card No.	Problems										
	Heterosexual conflicts	Homosexual conflicts	Sex triangle	Conflict with authority	Use made of power	Situational hope	Situational fate	Ethical standards	Personal adequacy	Alternative courses of action	None indicated
1.	—	—	—	44.9	—	16.6	2.6	2.6	14.1	7.7	11.5
2.	2.6	—	6.4	15.4	—	44.9	1.3	—	2.6	15.4	11.5
4.	62.8	3.8	20.5	1.3	—	7.7	—	1.3	1.3	—	1.3
5.	7.7	1.3	3.8	11.5	12.8	23.1	1.3	3.8	—	—	34.6
10.	14.1	3.8	1.3	1.3	—	19.2	19.2	1.3	5.1	—	34.6
11.	—	1.3	1.3	10.3	23.1	32.1	5.1	—	2.6	1.3	23.1
13.	21.8	—	5.1	1.3	1.3	7.7	24.4	25.6	5.1	3.8	3.8
14.	2.6	—	—	6.4	1.3	34.6	—	—	9.0	6.4	39.7
15.	6.4	—	—	3.8	5.1	14.1	20.5	14.1	7.7	—	28.2
16.	1.3	1.3	2.6	3.8	2.6	24.4	3.8	3.8	10.3	—	46.2
19.	1.3	1.3	1.3	6.4	1.3	30.8	2.6	—	2.6	—	52.6
20.	14.1	—	2.6	5.1	2.6	39.7	1.3	3.8	10.3	9.0	11.5

*The rows add to 100.0± a small error due to rounding of figures.

TABLE 3

Variations in Behavioral Response Patterns Induced by Different *TAT* Cards Shown by the Distribution of Frequencies Expressed as per cent*

(78 patients)

Behavioral response	Card Number 1	2	4	5	10	11	13	14	15	16	19	20
Aggressive Attack	1.3	—	2.6	3.8	1.3	26.9	15.4	2.6	5.1	6.4	—	11.5
Retribution or Revenge	—	—	2.6	2.6	—	—	1.3	1.3	3.8	2.6	1.3	—
Constructive Aggression	23.1	41.0	19.2	7.7	5.1	10.3	10.3	12.8	7.7	6.4	3.8	5.1
Rational Approach	2.6	12.8	28.2	7.7	6.4	1.3	2.6	7.7	6.4	3.8	5.1	10.3
Sensuous Gratification	—	—	1.3	3.8	1.3	—	5.1	—	—	1.3	1.3	—
Compromise	1.3	—	—	1.3	—	1.3	—	—	—	—	—	—
Seeks Help	7.7	—	2.6	1.3	14.1	2.6	3.8	2.6	7.7	3.8	6.4	9.0
Passive Hostility	6.4	1.3	1.3	1.3	—	—	2.6	1.3	2.6	—	1.3	1.3
Anxious Suspension of Activity	3.8	—	—	2.6	—	1.3	1.3	—	—	2.6	—	1.3
Acceptance of Situation	2.6	14.1	1.3	20.5	24.4	5.1	7.7	11.5	9.0	14.1	9.0	10.3
Irrational, Ill-directed Activity	—	—	1.3	1.3	—	1.3	1.3	—	1.3	—	1.3	1.3
Submission	16.6	5.1	9.0	7.7	3.8	3.8	7.7	2.6	1.3	1.3	3.8	2.6
Avoidance	—	1.3	2.6	1.3	1.3	—	1.3	—	—	1.3	—	1.3
Resignation	7.7	7.7	5.1	9.0	12.8	3.8	16.6	11.5	16.6	10.3	2.6	20.5
Active Withdrawal	11.5	5.1	20.5	7.7	5.1	21.8	9.0	6.4	5.1	5.1	2.6	7.7
Suicide	—	—	—	2.6	—	—	6.4	1.3	2.6	2.6	1.3	2.6
No Action	12.8	9.0	1.3	17.9	24.4	20.5	6.4	38.5	29.5	37.2	60.3	10.3
Alternative Actions (Unresolved)	2.6	2.6	1.3	—	—	—	1.3	—	1.3	1.3	—	5.1

*The columns add to 100.0± a small error due to rounding of figures.

the characters tends to be strong and antithetical—either some form of aggressive attack on the situation or flight from the existing pressures.

It is already evident from Table 1 that alcoholics do not differ from the assorted patient group in the kinds of themas elicited by the pictures common to all groups. When the data of that table are combined so that the alcoholics (both males and females) are compared to all the assorted patients, the chi square for problem-frequencies is 7.70. In order that the distribution of problem frequencies be significantly different at the 5 per cent level, the chi square would have to be 18.31. These figures indicate that alcoholics and assorted patients act like members of the same population in their selection of themes about which to weave a story.

Since no group differences in thematic emphasis were observed between alcoholic and assorted patients, we turn to an examination of the behaviors through which the patients moved their characters. As will be remembered, the upper seven items on the left of the check sheet represent a positive approach to situations ranging all the way from strong aggression to the mildly positive action of seeking help. The middle group represents a sort of neutral response wherein the story character makes no real effort to move into the situation, but neither does he tend to evade the problem. The lower group represents a negative approach in the sense of trying to avoid the situation with greater or lesser energy.

Because sex differences were observed on some of the common cards in terms of actions of the characters, the data were kept separate for behavior-responses to an individual card, and a compilation covering all 12 cards was made for the purpose of group comparisons. Taking the latter first and referring to Figure 3, it is seen that in progressing from assorted males, assorted females, alcoholic males, to alcoholic females there is an increase in the frequency of positive responses, a decrease in negative responses, and a slight increase in neutral responses. The chi squares between any two of these groups are significant at the 2 per cent level except for that between alcoholic males and females. Thus, alcoholics tend to tell more assertive stories then do other hospitalized patients.

The action frequencies most noticeably increased for the alcoholic patients are those in which the story characters attempt to solve the problems constructively or attempt to deal rationally with the situation. There is also a tendency for story characters to accept situations as they are. The assorted patients, on the other hand, show a significantly greater incidence of story actions such as submission, resignation, or withdrawal, and a greater percentage of stories involving no action. While the group differences shown

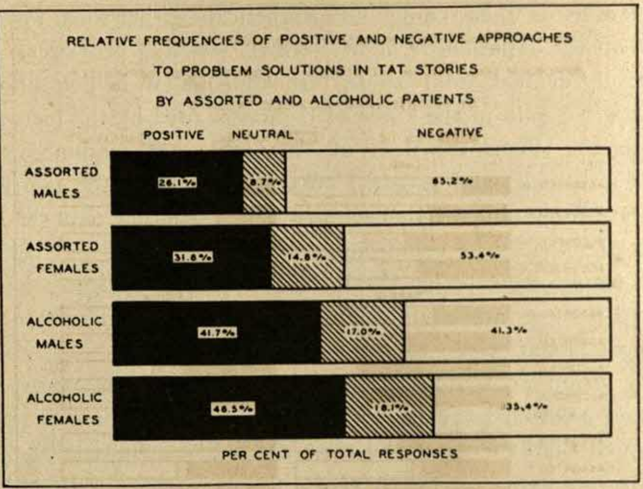


FIGURE 3

above are significant in a statistical sense, they represent only trends to which there are a large number of individual exceptions in both patient groups.

The action frequencies for each card separately in terms of the same three classifications are shown in the bar charts of Figure 4. Aside from the variations between patient groups in the actions of the story characters, an over-all consideration of the data of each card indicates that some cards seem to generate a greater frequency of positive responses in all patients than others. The greatest frequencies of positive responses were given to Cards 2 and 4. The lowest number of positive responses were to Card 19. Note that Card 4 elicits very few neutral actions for any patient group, whereas neutral actions are much more frequent for Cards 5 and 10.

The differential frequencies between the patient groups vary from card to card. As stated before, the alcoholic patients tell stories with somewhat more positive and many more neutral actions than assorted patients. In addition, female patients show a slightly higher incidence of positive action and fewer neutral responses to Card 5 than males. Proportional variations by sex are also seen in other cards, e.g., Cards 14 and 16.

These examples show that while there are little or no sex or group differences in the frequencies of plots or problems produced in stories for the common cards, the variations in actions obtained are related to both factors of sex and the characteristics of the patient group studied. It could probably be safely stated that actions ascribed to the characters in the *TAT* pictures

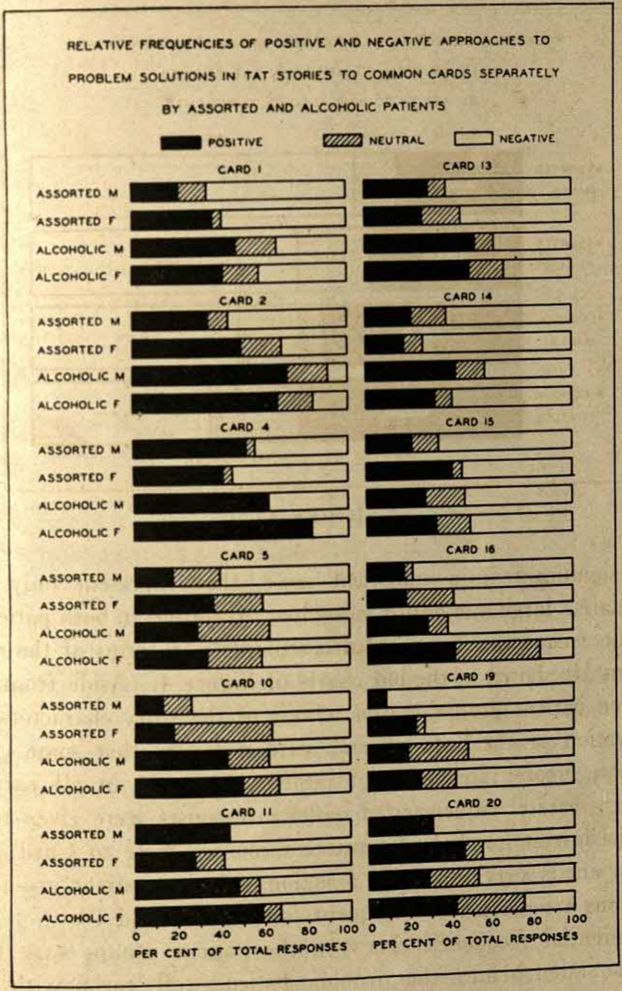


FIGURE 4

are a matter of individual personality, influenced to an unknown degree by the stimulus picture, the pathology, and the sex rôle.

In addition to variations in problem-action responses, we investigated differences in the emotional content expressed in the stories of the alcoholic and the other hospitalized patients.

The frequencies with which the patients ascribed emotional responses to the story characters are shown in Figure 5. It will be observed that there are only slight differences in frequency of emotions between assorted males,

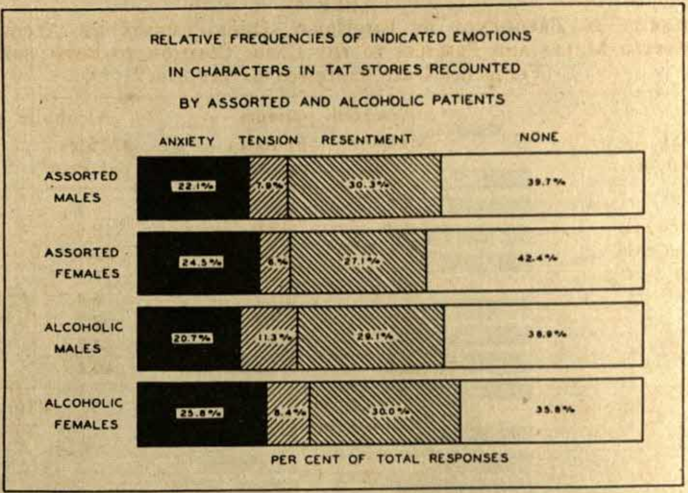


FIGURE 5

assorted females, alcoholic males, and alcoholic females. The chi square between the assorted males and females is .66. (*P* at .05 equals 7.82.) The chi square between alcoholic males and females is 1.11. Since the chi squares indicate sampling from the same population, the data for the males and females of assorted and alcoholic patients respectively, were combined. The chi square between all alcoholics and all assorted patients is then .79 with a *P* at .05 as above. These data indicate that the alcoholics attribute neither a significantly greater or lesser degree of anxiety, tension, or resentment to their story characters than do the nonalcoholic patients.

In a comparison of the relative frequency of the situations or persons to which indicated emotional involvement was referred, we again find (Table 4) there is an absence of significant group differences between the assorted males and females and alcoholic males and females. The chi square between the latter is 4.10 (*P* at .05 equals 14.07), whereas it is 2.55 between the assorted males and females. When the data for males and females are combined for each of the larger clinical groups, the chi square is 1.55.

Two of the small differences in Table 4 are interesting. When compared to other patients, alcoholic males show a smaller frequency of emotional response referred to "underlings." Assorted females show somewhat less emotional displacement to "situations" than do the other patients.

Since some of the frequencies for emotional displacement are small, chi squares calculated on these data could be criticized. Correlations of the

TABLE 4

THE SIMILARITY IN FREQUENCY OF EMOTIONAL DISPLACEMENTS BY ALCOHOLIC AND ASSORTED MALES AND FEMALES TO THE CARDS COMMON TO BOTH SEXES
(Frequencies are expressed as per cent)

Emotional displacement	Assorted patients		Alcoholic patients	
	Males per cent	Females per cent	Males per cent	Females per cent
Self	8.2	7.7	8.1	7.2
Same sex-equals	3.2	2.0	2.6	4.2
Other sex-equals	8.5	12.0	12.4	12.6
Underlings	4.1	4.7	1.7	6.0
Authority	6.9	7.7	6.4	7.2
Society	3.2	3.7	3.3	1.2
Situation	23.3	15.7	24.7	25.1
None	42.6	46.5	40.8	36.5
Chi Square	2.55		4.10	
P	> .90		> .70	

data (raw data) for alcoholic males and females, and again for assorted males and females gave *r*'s of .99 and .97 respectively. This is another way of saying that the relative frequencies of emotional displacements to the headings shown by the males in each group could be predicted with fairly high accuracy by knowing the same data for the female groups. This provides a check on the chi squares and indicates that these groups (male and female) do not differ significantly in the frequencies of their characters' emotional displacement on the common cards of the *TAT*. A similar result is obtained between the total groups of alcoholic and assorted patients with an *r* of .98.

As in the case of problem-action responses, there is a tendency for some cards to produce a greater incidence of emotions in the story characters than others. Here the correspondence between males and females of each clinical group is slightly less marked in some categories. Figure 6 shows the relative incidence of emotional involvement attributed to story characters in the common cards. Since the values are expressed as frequency per patient, an idea of the proportion of patients in a given group expressing a particular emotional response may be gained. For example, a value of .5 means that, on the average, one out of every two patients would attribute the emotion in some strength to the story characters involved in the particular card.

When male and female patients in both clinical groups are compared within each group, it is seen that the corresponding patterns are remarkably alike in resentment and less similar in anxiety and tension. Data already shown above indicate that there is a slightly greater but not significant

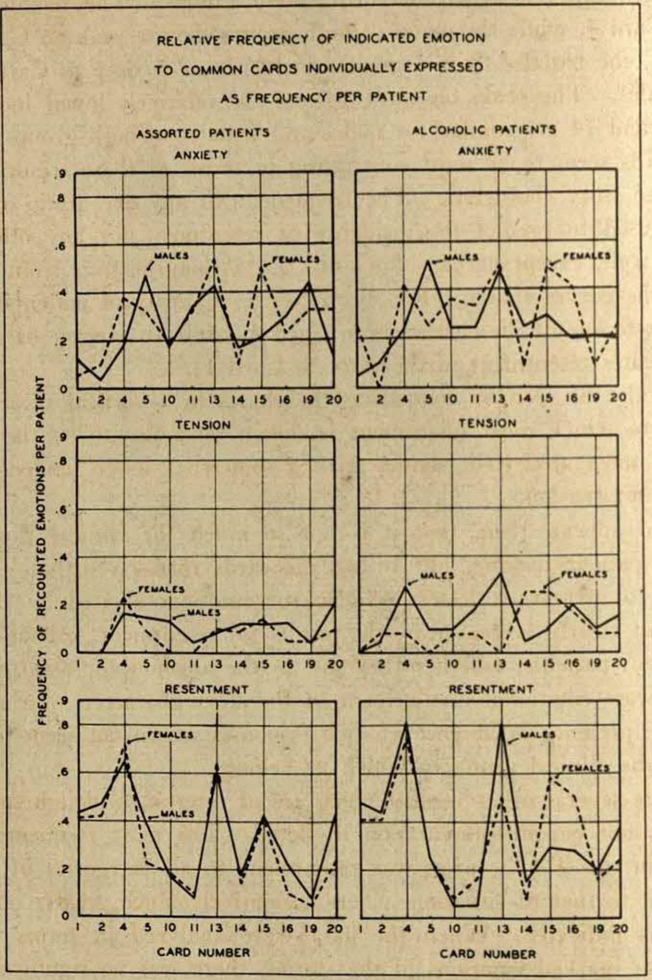


FIGURE 6

recounting of tension by alcoholics. There was generally little expression of tension in the story characters. No card elicited a tension response from more than one patient in three, while no patient in any group indicated tension in Card 1. While the tension patterns are more alike for assorted than alcoholic patients, the incidence is so low that no account can be given for the better agreement in assorted patients.

In the case of anxiety, the different cards tend to call for similar frequencies in the alcoholic and assorted groups. Here, however, there are apparent sex differences within both patient groups. The females show a peak on Card 4, while the males of both groups show a peak on Card 5. In both cases, the females show a greater incidence of anxiety to Card 15 than do the males. The peaks on Card 13 and the relatively lower incidence on Cards 10 and 14 are similar for males and females of both groups.

The cards seem to be most controlling in terms of the attribution of resentment to story characters. The frequencies of any one group of patients could be used to predict the incidence of resentment for any other group relatively well, except perhaps for Card 15, as may be seen from the similarity of the curves shown. For example, the majority of patients attribute resentment to one of the characters in Card 4, while an average of about one in 10 indicates resentment in the story to Card 11.

It may also be seen, by comparing the charts in a vertical direction that the cards to which most resentment is shown tend also to be the cards to which the most anxiety is shown with a somewhat lesser tendency of the same kind for tension.

It would appear, then, that it is not so much the clinical grouping to which the patients belong, but rather the cards themselves that determine, in general, a greater or lesser incidence of emotions elicited by the cards. This is most clearly evidenced by the results for resentment. About the same proportion of patients, whether they fall in the alcoholic or assorted group, fail to indicate emotional involvement of the story characters. In any given patient, the presence or absence of such responses would not seem to be predictive of the clinical group to which he belongs.

There are several other results which are of interest, although space limitation does not permit presentation in detail. The most frequent emotion recounted in the *TAT* stories was resentment to all categories of displacement except to that of situation, where a relatively much greater amount of anxiety was indicated. When the data were analyzed in terms of mild, moderate, or marked emotions in the stories, there was no significant trend in either the alcoholic or assorted patients as to the strength of the emotions indicated for the story characters.

Biochemical studies (to be published elsewhere) were made on 18 alcoholics being reported on in the present study. The *TAT* protocols were obtained at about the same time the laboratory tests were done. In 15 of these patients the two sets of data were only several days apart, whereas for the other three the spacing ranged from 19 to 28 days. Correlations were obtained between the incidence of the emotions in the *TAT* as scores weighted in terms of intensity and the corresponding determinations of biochemical substances measured in gamma. The correlations for anxiety, tension, and resentment were .28, .40, and —.44 respectively. The *t* ratios for these coefficients fail to reach the 5 per cent level. Since we cannot reject the null hypothesis with confidence, we cannot say that our correlations have not arisen by chance.

In general, our abstraction of the emotions of anxiety, tension, and resentment from the *TAT* stories has not yielded any differentiating characteristics which would be more descriptive of the alcoholic's response to the test than that of other hospitalized psychiatric patients.

D. Discussion

We have been aware that there are real difficulties involved in a comparison such as we have made between personality dynamics revealed in clinical treatment and those extracted by the *TAT* stories.

Our findings that the *TAT* has not contributed to the understanding of more than one patient must be viewed in terms of the situation in which the comparisons have been made. The physicians were generally on a given case for about three weeks prior to testing and the time spent with each patient was relatively great. Therefore, there was adequate opportunity to study the etiological factors in the pathology. It is possible that some factors crystallized by the *TAT* might have been previously known to the physician as a vague clinical impression. In the case of the one patient mentioned, we cannot be sure that a few more interviews with the patient would not have brought forth the new dynamic material.

Since this study has been carried out in a teaching hospital, the day to day clinical usefulness of the procedure represents a problem worthy of evaluation. The failure to uncover personality dynamics beyond those ordinarily observed relatively early in the course of therapy contraindicates its use in routine clinical evaluation. Even in those cases where the test has been used hopefully in attempting to break through resistance or blocking in clincial interviews, the results have been disappointing. The *TAT* is a time-consuming procedure at best. In our experience, six to eight hours of a psychologist's time added to that of the physician usually yields no more diagnostic information than the physician's interviews alone. There seems, then, little point in its frequent use in the circumstances peculiar to us.

A major problem concerning this or any other test device is that of validation. The test evaluation of the isolated individual becomes critical where general test validation is not easily established. In the case of the *TAT* wherein subjective evaluation is the rule, one cannot be sure that the personality dynamics gleaned from the stories represent significant statements about the personality characteristics of the raconteur. The choices here are left up to the examiner. By what criteria, for example, does he state that certain characteristics of the hero represent wish-fulfillment in the patient, or that this is analogous or that, complementary? Clinical insight is an obvious aid, but then, with adequate clinical insight one need not lean on the test procedure.

There are others who question the value of the *TAT* as an instrument of diagnostic aid, both in terms of the economy of procedure and its capacity to discriminate personality trends in patients. In a review of psychodiagnostic instruments, Challman (2) feels that recent studies throw some question on the present widespread use of the test as a clinical tool. Sanford's remarks in the foreword to Aron's *Manual for the Analysis of the Thematic Apperception Test* (1) suggests that the practical usefulness of the test is no better than when first published. Among other things, he states that the *TAT* does not premit predictions of sufficient "probability" that he is willing to make decisions upon them which affect the lives of patients. He considers the test a luxury in the clinical situation where the examiner might do even better if able to conduct primary interviews. We have come to similar conclusions in this paper.

The basic requirement of useful tests or techniques developed in any scientific area is that they aid in making more precise, consistent, and detailed determinations of the phenomena studied than are obtainable by commonsense observations alone. Any test to be of value must satisfy this criterion, or its use must be rationalized on some other grounds. It is obvious from studies on the *TAT* reported in the literature and from our clinical experience with this and other projective techniques that, while the *TAT* in some instances may give a fairly good picture of the dynamics of the patient's psychopathology, the test fails to satisfy the above criterion. That is, it does not appear to have the characteristic of giving us more reliable and precise determinations of motivational variables, for example, than usual reasonable clinical insight. One uses scales in weighing substances since the procedure has the value of eliminating much of the error in the Weber fraction. Certain aptitude tests have the characteristic of making psychologists' predictive estimates more reliable. We may detect unaided certain hearing losses in an individual, but audiometric devices give a much clearer indication of the pattern and extent of the loss. In the present instance, there is a very real question as to how much the errors of clinical observation are reduced by the use of a projective device such as the one here under consideration.

It is characteristic of projective tests that they tend to vary with respect to the degree of structuring which they exhibit. It is probably generally conceded that the *TAT* is more structured than the Rorschach Test. However, structuring must be only relative, since the only truly unstructured test would probably be one in which the individual experienced an homogeneous stimulus pattern, in which case any variability of response is attributable to characteristics of the organism itself. The moment some clue is introduced into the situation, a certain amount of structuring must be admitted, for the response is now determined by both sensory data and central processes and by neither alone.

From the literature one gains the impression that the *TAT* cards lend themselves to more variable interpretations than we have found in this study. An examination of the frequency of plot-responses on the different cards indicates that each card tends to produce a pattern of its own with respect to the nature of the problem which the characters face in the stories. Each card also produces its own pattern of relative frequencies with which anxiety, tension, and resentment are attributed to the story characters. Similar findings have been reported by Garfield and Eron (8) who indicate that the themas or moods appear to be a function of the stimulus values of the cards used. Eron (6) has also shown in another study that the types of themas produced by schizophrenic patients and college students appeared to be relatively more a function of the stimulus material than of personality characteristics of the individuals.

In a study similar in some respects to the present one, Eron (5) showed that fantasy productions in terms of themes and outcomes did not differ significantly between neurotic and schizophrenic patients, and these in turn did not differ from an equated group of normals. He concluded that the *TAT* cannot be used as a diagnostic instrument in the sense of its yielding signs or patterns characteristic of specific nosological groups. In this study we have been unable to distinguish between alcoholics and other patients in terms of the patterns of problem frequencies.

Thus, while the results are by no means conclusive, a serious question is raised regarding the "projective" aspect of the themas of the *TAT*. It is apparent from the data that some freedom is shown by individuals. But by which individuals, and to which cards and under what circumstances is the test "projective"? The greater freedom in what the characters do once the scene is set seems to be the more projective part of the test. Should this be a statement of relative fact, there still remains the problem of meaning. There always seem to be stories told in lieu of the first one thought of and multiple possibilities of interpretation. These results again point up the importance of validating studies.

The overall tendency for alcoholic patients to tell fewer stories involving submission and retreat from problems is interesting in the light of the popular notion that alcoholism represents some form of escape mechanism. There may be other factors which would account for the difference between the alcoholic and assorted patients, such as the greater incidence of schizophrenic patients in the assorted group, but it would be interesting to determine how far the outgoing stories of the alcoholic patients represent a compensatory characteristic. In a sense, the alcoholic is positively engaged in doing something about his troubles albeit an illdirected activity of drinking. These results are at variance with a study done by Klebanoff (9) wherein the alcoholic shows a relative lack of aggressive tendencies and much emotional tension related to themas involving internalized conflict. It would appear that our results are almost the reverse of Klebanoff's findings. We suspect that the trends observed are more a function of associated pathology rather than the factors which induced alcoholism in the particular patient.

So far as the problem of the alcoholic and his emotional problems are concerned, it has been suggested by clinical and biochemical studies (7) that in drinking, the alcoholic finds a means of reducing resentment, anxiety, and tension, which appear to be perhaps more frequent in alcoholism than in other nosological groupings. Our interest here was the possibility of finding different patterns of emotional involvement in the alcoholic patients in terms of their productions using the *TAT* pictures. The negative results reported above indicate that the patterns of the alcoholic and assorted patient groups are similar both in terms of incidence and distribution of the three emotions discussed, and in the persons or situations to which the emotions were referred. There are at least two possible interpretations of these results, and we are without adequate basis for deciding between them. Either alcoholics in general do not differ from other patients insofar as emotional patterns are concerned, or we have failed to discern differences that are peculiar to alcoholics when compared to other patients, due to the limitations of the *TAT* technique, our present application of it, or both.

We are well aware that our approach to an individual analysis of a *TAT* protocol is different from our more objective approach in this study, particularly with regard to the problem of emotion. For example, an entire story might imply but not state a feeling of tension, yet in the analyses as we have made here, one had no basis for saying that a particular story character was even mildly tense. These impressions which can be integrated into an individual report must be sacrificed in some circumstances in the interest of an approach to objectivity. As we stated earlier, generalization is usually made at the expense of specificity, and it is to generalization that we have turned in this study.

As was shown with respect to problem-action analyses, we have found that the incidence of some emotions to particular cards could be fairly accurately predicted in patient groups. This was particularly true in the case of resentment. Again we see that some cards provide relatively less freedom for "projection" of emotions. Therefore, the presence of emotional involvement of this kind in these cards indicates perhaps more about the picture than about the patient.

In general, by analytical treatment of the *TAT* stories, we have not observed specific personality dynamics which set alcoholic patients apart from other hospitalized patients. This result is in agreement with only one of two published studies known to us, in which the *TAT* has been used to determine possible personality characteristics which would identify alcoholics. Klebanoff (9) reports characteristics which we do not find. Further, the differences obtained in the present study may be referable to factors other than trait patterns of alcoholics per se. The other study by Roe (12) compared artists who were moderate, social, and excessive drinkers. It was found that one could not pick out the excessive drinker by blind analysis of *TAT* protocols.

Returning to the present study, it may well be that our experimental approach has been adequate but revealed no personality characteristics in alcoholics as a group simply because these patients differ among themselves as much as they differ from other patients. This returns us to the fundamental problem of why some individuals become alcoholic and others do not. The evidence in this study sheds little light on the matter. Yet the search for the determiners of such behavior should not lapse, even if psychological study should ultimately show that alcoholism was the end result of fortuitous events in the life of the individual related perhaps to body chemistry, and that alcoholism could be predicted with only chance success prior to exposure to alcohol in any form.

E. SUMMARY

The personality dynamics of 84 inpatients inferred from *TAT* protocols were compared with the physicians' knowledge of personality dynamics of the same patients within the first month of hospitalization. New but relatively unimportant dynamics were revealed by the analysis of the *TAT*'s in 13 cases, while clinically significant new material was revealed for only one patient.

The test did not generally provide information beyond that obtained in the usual clinical therapeutic sessions. In many instances it provided little or no information of diagnostic or clinical significance. It is not considered to be an economical or efficient procedure in a situation where clinicians are well-trained and the patient load is such that intensive study of the patients is the rule. Where the physician was blocked by the patient's defenses, it was hoped that the test would provide a means of getting through to important dynamics. Ordinarily we have not penetrated such resistances.

The *TAT* appears to be more structured as to the problems or themas induced by the pictures, with different cards showing different patterns of problem frequencies. Sex differences are not significant. Less structuring is observed with reference to the responses or actions of the story characters in dealing with the problems they face.

Comparison of the total alcoholic patient group with all nonalcoholic patients in terms of problem-action frequencies for the 12 common cards combined, showed no difference in the frequencies of the problems selected, but yielded significant differences as to the nature of the actions attributed to the story characters. The tendency is for the alcoholic to tell stories with outgoing, positive action in contrast to nonalcoholic patients, who tell a greater number of stories in which the characters are submissive and avoid the problems. We are inclined to account for the difference in terms of associated psychopathology. Also, since there are many individual exceptions to the trend noted, such information could not be used in the analysis of individual protocols.

By our method of analysis, no significant differences were observed between the patient groups in the incidence of the emotions of anxiety, tension, and resentment ascribed to the characters of the *TAT* stories. Nor was a difference found between the clinical groups in the relative frequencies of emotional displacements. Thus, as a group, alcoholics are projectively like nonalcoholic psychiatric patients in their emotional responses to the *TAT* pictures.

The data provide no evidence of personality characteristics common to alcoholics as a group which distinguish them from other patients.

REFERENCES

1. ARON, B. A Manual for Analysis of the Thematic Apperception Test. Berkeley, Calif.: Berg, 1949.
2. CHALLMAN, R. C. Clinical methods: psychodiagnostics. In C. P. Stone (*Ed.*), *Annual Review of Psychology*, 1951.
3. COX, B., & SARGENT, H. *TAT* responses of emotionally disturbed and emotionally stable children: Clinical judgment vs. normative data. *J. Proj. Tech.*, 1950, **14**, 60-74.

4. DIETHELM, O., & KNEHR, C. A. The diagnostic use of psychological tests from the psychiatrists' standpoint. In P. H. Hoch (*Ed.*), *Relation of Psychological Tests to Psychiatry*. New York: Grune & Stratton, 1952.

5. ERON, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monog.*, 1950, **64**, No. 9 (Whole No. 315).

6. ————. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. Consult. Psychol.*, 1948, **12**, 387-395.

7. FLEETWOOD, M. F., & DIETHELM, O. Emotions and biochemical findings in alcoholism. *Amer. J. Psychiat.*, 1951, **108**, 433-438.

8. GARFIELD, S. L., & ERON, L. D. Interpreting mood and activity in *TAT* stories. *J. Abn. & Soc. Psychol.*, 1948, **43**, 338-345.

9. KLEBANOFF, S. G. Personality factors in symptomatic chronic alcoholism as indicated by the Thematic Apperception Test. *J. Consult. Psychol.*, 1947, **11**, 111-119.

10. MURRAY, H. A. Thematic Apperception Test Manual. Cambridge: Harvard Univ. Press, 1943.

11. ————. Uses of the Thematic Apperception Test. *Amer. J. Psychiat.*, 1951, **107**, 577-581.

12. ROE, A. Alcohol and creative work: Part 1. Painters. *Quart. J. Stud. Alcohol*, 1946, **6**, 415-467.

13. ROTTER, J. B. Thematic Apperception Tests: Suggestions for administration and interpretation. *J. Personal.*, 1946, **15**, 70-92.

14. TOMKINS, S. S. The present status of the Thematic Apperception Test. *Amer. J. Orthopsychiat.*, 1949, **19**, 358-362.

Payne Whitney Psychiatric Clinic
The New York Hospital
525 East 68 Street
New York City 21

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 227-228.

A VISUAL IMAGERY TECHNIQUE FOR THE INDUCTION OF HYPNOSIS IN CERTAIN REFRACTORY SUBJECTS*

Department of Psychology, Long Island University

MILTON V. KLINE

A. INTRODUCTION

Contemporary advancements in hypnotherapeutic techniques and applications have again focused attention on the problems of hypnotic induction, hypnotic resistance, and hypnotic depth. At present a wide variety of induction techniques exist which permit a majority of patients in psychotherapy to utilize hypnotic procedures both in directive and non-directive treatment situations. Clinical therapeutic experience has indicated that in many cases of psychotherapy a light hypnotic state is not only adequate but often is more desirable than a "deep" trance state, particularly in intensive therapy. Nevertheless, there are a considerable number of patients with whom a hypnotherapeutic approach would be indicated except for their resistance to induction. In some of these cases, the use of a good deal of additional time in the induction phase will sometimes yield positive results, but such a use of treatment time is not often practical.

The purpose of this paper is to describe an hypnotic induction technique which has proven effective in inducing light to medium trance states rapidly in patients otherwise refractory to hypnotic induction. All of the subjects involved in this report gave evidence of being able to use the hypnotic state productively with respect to therapeutic techniques.

B. TECHNIQUE

The subjects reported on here were those with whom hypnosis could not be achieved with a variety of techniques and the expenditure of not less than 20 minutes. The visual imagery technique employed with this group was as follows:

1. In the waking state with eyes open, each subject was asked to visualize in "his mind's eye" certain familiar objects. In order these were: (*a*) a house, (*b*) a tree, (*c*) a person, and (*d*) an animal. The psychodiagnostic value of this imagery production will be dealt with elsewhere. This step was continued until each stimulus had been achieved. In this population of 15 subjects, all were able to achieve the requested images readily and easily.

---

*Received in the Editorial Office on October 13, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

For subjects who may have difficulty in visual imagery, other methods may have to be devised based upon the principles described here.

2. Following the attainment of image formation in the waking state, each subject was told, "Close your eyes and in your mind's eye visualize yourself as you are here; sitting in the chair (or lying on the couch) *except the image of yourself has his (her) eyes open.*"

3. At this point the subject was told to concentrate on the image and that all the therapist's (experimenter's) comments would be directed toward the subject's *image* and *not* toward the subject.

4. Then, a simple ocular-fixation technique was described and related to the eye-closure of the *image.* Close clinical observation of the subject will reveal subtle response patterns indicating the associative effect upon him directly. The subject can be asked to confirm eye-closure in the image, though often his own straining to raise his eyebrows will reveal the situation. The image can be challenged on the lid catalepsy depending on the value that this mechanism may have in the total hypnotic relationship. Following eye-closure in the *image,* suggestions for "deepening" the trance are given in the usual manner.

5. The next step involves moving directly into the induction relationship with the *subject.* This may be done by saying, "Now you are feeling just like the image, going deeper and deeper asleep (or an equated word) and the image is disappearing." Within a few minutes, depending on the subjects' personality, you will have obtained a light to medium hypnotic trance. Further depth may be secured in the usual manner, but the patient is now ready for hypnotherapeutic work.

The total elapsed time involved in this induction technique has averaged 10 minutes. There is more deviation on the less time side than the more time side quantitatively. Utilizing post-hypnotic methods, subsequent inductions have usually been obtained directly with one of the standard methods. Although limited to 15 subjects, this visual imagery technique has enabled all 15 subjects to achieve an hypnotic state for which they had been refractory with other methods. This obtained state was adequate for hypnotherapy.

C. SUMMARY

This paper has described a visual imagery technique for the induction of light to medium states of hypnosis in refractory subjects. Though limited to a small sampling of such subjects, this technique has proven effective, rapid, and easy to manage within the framework of the hypnotherapeutic relationship. Further application of this technique appears warranted.

500 Riverdale Avenue
Yonkers, New York

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 229-233.

AN ACADEMIC APTITUDE TEST FOR SUPERIOR COLLEGE STUDENTS*

Department of Psychology, San Francisco State College

H. C. Lindgren, R. Gilberg, and D. W. Crosby

A. Purpose

The purpose of this study is to report preliminary data resulting from an investigation of the validity and reliability of a proposed measure of academic aptitude for superior college students.

B. Background and Method

The earliest form of the Linguistic Aptitude Test (hereinafter referred to as the *LAT*) was constructed by the senior author, who was in need of a brief measure of academic aptitude for superior students requesting vocational and educational guidance in a college guidance center. He found no lack of aptitude tests appropriate for use with college freshmen, tests which were quite adequate for purposes of counseling students who were interested in completing the standard undergraduate curricula. However, he found that there are few tests available to the counselor who has the task of advising the student who is interested in working for a graduate degree. Thorndike's *CAVD* is lengthy and difficult to score; the Miller Analogies Test is available for use only under rigidly controlled conditions; and the Graduate Record Examinations cannot be secured and administered quickly and cheaply. As a matter of fact, neither of the latter two tests were available at the institution in question.

It therefore seemed that what was needed was a test which would be brief, cheap, easily administered, and easily scored, and which would measure, with some degree of validity, some of the important dimensions of whatever it is that students need to have in order to succeed in graduate work. For the purposes of the problem in question it was decided that such a test should be linguistic in nature, because of the importance of the linguistic factor in all graduate work, and because most of the advisees of the senior author were primarily interested in graduate work in the fields of the liberal arts and the social sciences. In order that the proposed test should reproduce some of the

---

*Received in the Editorial Office on October 16, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

conditions of graduate work, it was also determined that the problems should possess a high level of real or apparent difficulty.

In other words, graduate work was envisioned, for the purpose of this test, as consisting of constellations of linguistic problems involving a high degree of frustration. Because the senior author was familiar with problems of foreign language teaching, it seemed to him that the task of trying to evolve some meaning from passages written in an unfamiliar but cognate foreign language contained in essence the elements of linguistic abstraction, difficulty, and frustration which characterizes much graduate work.

Accordingly, a test was constructed consisting of passages written in Dutch and in Early Middle English, each of which was followed by multiple-choice questions in Modern English covering information and concepts presented. No further aid, in the way of clues or hints, was given.

It was believed by the senior author that this type of test would be more effective than the artificial language tests employed as subtests by some authors of academic aptitude tests, because the etymological relationship between these two languages and English meant that the subject would in each case start with some advantage. Yet it would be an advantage which would not be apparent to the subject at first glance, but which would become more useful to him as he proceeded with the problem of extracting meaning from the passages in question. On the other hand, if the subject were easily frustrated by complex and abstract linguistic situations, he would be less likely to perceive the relationship between the unfamiliar languages and Modern English, and consequently would not be able to score very high on the test.

This initial form of the *LAT* was put through two revisions, based on item analyses. In its present form, it takes 45 minutes and consists of 36 questions based on three passages of Dutch and two passages of Early Middle English.

C. VALIDATION

The present study of the *LAT* is based on its administration to some 602 undergraduate students at San Francisco State College, a liberal arts college in a municipal setting. Subjects were obtained by administering the *LAT* to classes whose instructors made them available for this study. It would have been more desirable to have used graduate and near-graduate students to serve as the base population for this study; however, only undergraduates were available for study (most of them freshmen), and it was decided that while data based on these populations would leave unanswered the question of the *LAT*'s ability to predict success in graduate work, they would provide valuable information with regard to the general utility of the test. In other words, if it were possible to demonstrate some validity with regard to undergraduates, this would provide encouragement for further studies with graduate students when populations of the latter type became available.

The corrected split-half reliability of the *LAT* was found to be .79 for seniors ($N = 49$) and .57 for freshmen ($N = 200$). (This is not surprising in view of the fact that the test was not designed for freshmen.) However, when the reliability is computed for freshmen above the 75th percentile on the American Council on Education Psychological Examination for College Freshmen, (hereinafter referred to as the *ACE*)—in other words, the superior student—the corrected split-half reliability rises to .66 ($N = 65$). It is .75 for all nonfreshmen ($N = 51$) who are above the 75th percentile on the *ACE* or the Ohio State University Psychological Test (hereinafter referred to as the *OSU*). Although higher reliability coefficients would be more desirable, these data indicate that the *LAT* is reliable enough for limited or experimental use with populations of superior or advanced students. It is possible that its reliability is high enough for use with graduate students, although evidence on this point is lacking.

There is a relatively high relationship between the *LAT* and other College aptitude tests. It correlates as follows with the *ACE*: Seniors: .60 (12);[1] Juniors: .77 (20); Sophomores: .71 (24); nonfreshmen: .71 (56); freshmen: .39 (363). The correlation between the *LAT* and the *OSU* for 60 nonfreshmen was .62, and the correlation between the *LAT* and the Diagnostic Reading Test for 335 freshmen was .34.

The investigators were also interested in exploring the relationship between previous training in foreign languages and scores received on the *LAT*.

TABLE 1

MEASURES OF CENTRAL TENDENCY AND DISPERSION FOR VARIOUS GROUPS OF UNDERGRADUATES TAKING THE LANGUAGE APTITUDE TEST

Group	Number	Mean	Median	Standard Deviation
Freshmen	388	15.09	14.92	3.91
Sophomores	36	17.00	17.70	5.28
Juniors	35	17.80	16.20	4.96
Seniors	27	18.59	18.38	4.92
Total for Nonfreshmen	101	17.60	17.43	5.11
Superior*—Nonfreshmen	42	20.79	20.33	4.72
Superior*—Freshmen	50	17.08	16.88	4.34

*Over 75 percentile on *ACE* or *OSU*.

---

[1] Figures in parenthesis refer to the numbers in each group.

Students who reported having studied foreign languages previously and those who reported no language were paired on the basis of scores received on the *ACE* and the Diagnostic Reading Test.

The difference of 3.0 between the mean score for the 26 persons who had studied German and the mean score for 26 who had studied no foreign language was significant at the 5 per cent level of confidence ($t = 2.75$). The mean difference of 2.26 was significant at the 1 per cent level for 43 persons who had studied two years of foreign language other than German and another 43 who had studied none ($t = 3.59$). But there was no significant difference between the means for similarly matched groups who had studied only one year of foreign language other than German (25 in each group; $t = —.14$); three years of foreign language (27 in each group; $t = .04$); four years of foreign language (25 in each group; $t = —1.48$); and five years or more of foreign language (13 in each group; $t = 1.39$). It appears as though German is of some advantage, probably because German is cognate with the languages which compose the test. Why *two* years of studying foreign languages other than German should appear to provide more of an advantage than three or more years is somewhat puzzling, unless there are characteristics which are peculiar to people who have completed the usual sequence of two years of foreign language in high school. At any rate, it will not be possible to ascribe any influence to previous foreign language training until further investigation has been made of these factors.

Table 2 indicates the relationship between grade point averages and various measures of academic aptitude, including the *LAT*. Inasmuch as the *LAT* was given in the spring of 1952, only the grade-point average for that semester

TABLE 2

CORRELATIONS BETWEEN GRADE-POINT AVERAGES AND VARIOUS MEASURES, INCLUDING THE LANGUAGE APTITUDE TEST

	Grade-Point Average, Spring Semester, 1952, only						Total Grade-Point Average as of June, 1952					
Group	*LA*	*N*	*ACE*	*N*	*OSU*	*N*	*LA*	*N*	*ACE*	*N*	*OSU*	*N*
Freshmen	.30**	303	.40**	301	***	—	.36**	301	.42**	301	***	—
Sophomores	.31	37	.66**	24	.31	15	.42**	40	.60**	24	.54*	16
Juniors	.44**	34	.53**	20	—.24	11	.40**	35	.47*	20	.37	15
Seniors	.52**	29	.47	12	.24	15	.56**	31	.22	12	.69**	15
Total non-freshmen	.41**	100	.57**	56	.31*	41	.46**	106	.50**	56	.57**	46

*Significant at 5 per cent level.

**Significant at 1 per cent level.

***None given.

can be considered as indicative of its predictive efficiency. However, correlations of the total grade-point averages are included for comparative purposes. It will be noted that the predictive efficiency of the *LAT* increases with each college class, with correlations ranging from .30 for freshmen to .52 for seniors. The lower "*r*" for freshmen is undoubtedly due, in part, to the lower reliability of the *LAT* for this group. Furthermore, the fact that the *LAT* correlates lower with grades than does the *ACE* or the *OSU* is very likely due to the lower reliability of the *LAT*, at least in part. Nevertheless, the data reported in Table 2 do point to the possibility that the *LAT* may have promise as a measure of academic aptitude for students who plan to undertake graduate work, particularly if its reliability is increased. The results with nonfreshmen are particularly encouraging.

D. SUMMARY

The need for a predictive measure of academic aptitude for students who are interested in undertaking graduate study led to the development of a Linguistic Aptitude Test consisting of passages in Dutch and Early Middle English, followed by multiple-choice questions in Modern English. The test was based on the hypothesis that a large portion of graduate work consists of dealing with constellations of highly abstract, difficult, and frustrating problems of a linguistic nature. The Linguistic Aptitude Test was administered to groups of undergraduates in an attempt to gather preliminary data respecting the general utility of the test. The results of these investigations indicate that although the reliability and validity of the *LAT* are not generally as high as most tests of academic aptitude, correlations of a respectable caliber were obtained with academic grades. Its efficiency as a predictor of academic success might very likely be improved if its reliability were increased. Further research is needed, particularly with graduate populations. The senior author invites correspondence with persons interested in the experimental use of the test in situations which will provide a test of its validity.

Department of Psychology
San Francisco State College
124 Buchanan Street
San Francisco 2, California

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 235-240.

A NOTE ON "FAMILIAR SIZE AND THE PERCEPTION OF DEPTH"*

Department of Psychology, Princeton University

WILLIAM H. ITTELSON

A. STATEMENT

In a recent article in this JOURNAL, Hochberg and Hochberg (3) state (*a*) that the only visual depth cue which necessitates the use of the explanatory concept of past experience is "familiar size,"[1] (*b*) that the evidence for the existence of such a cue is scanty and equivocal, and (*c*) that theoreticians who assign some importance to this cue are at best "premature."

The chief argument advanced in support of the second proposition (we shall not in this paper be concerned with the first or third) is that most of the evidence purporting to support the existence of a cue of familiar size actually involves relative sizes and that the effect of relative size can be accounted for without invoking past experience. Specifically, "it has yet to be shown that familiar size as opposed to *relative size* can be of any influence in 'primary', immediate space perception" (3, p. 113), and hence "we can hypothesize an autochthonous tendency toward homogeneity by which similar shapes 'seek' similar size through appropriate distribution in perceived space" (3, p. 109).

While these arguments are not new, they remain important because, if they should turn out to be in accord with the evidence, they would greatly advance our understanding of at least one part of the immensely complex and difficult study of space perception. Recognizing the great potential value of these arguments, the present author has attempted to apply them to various experiments of his own and of his colleagues. That this attempt has turned out to be a failure is undoubtedly due to his own shortcomings rather than to any inadequacies of the principles themselves. The purpose of this note is, therefore, to raise the question of how the following evidence, which is not

---

*Received in the Editorial Office on October 17, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The terms "familiar size" and "the cue of familiar size" will be used in this note since they are used in the paper under consideration and the introduction of new terms would be needlessly confusing. However, the author has elsewhere (4, 5) elaborated the preferred term of "assumed size" which has the advantages of being a much more general concept and at the same time being more amenable to rigid definition.

considered by the Hochbergs, can best be reconciled with the arguments advanced by them.

B. ARGUMENT

Five cases will be described. In the first three relative size is apparently *not* involved since only a single object is viewed at a time while in the remaining two cases an "autochthonous tendency toward homogeneity" does not appear to apply since exactly the same retinal pattern results in two different perceptions due to two different familiar sizes.

1. If a single familiar object is viewed monocularly in an otherwise dark room, it is perceived, immediately and unequivocally, at some definite distance, which can be correctly predicted on the basis of the familiar size of the object. Evidence for this statement is to be found in the original experiment reported by Ames (1) which has been replicated by several experimenters including Ittelson (4) and Lawrence (9). Although the Hochbergs do refer to these three publications, they do not explicitly consider this evidence which seems to the present author to offer at once the most elementary and the most conclusive demonstrations that familiar size can function as a cue to distance.

2. If a single, familiar object is viewed monocularly in an otherwise dark room, and then is replaced by another single, familiar object of exactly the same physical size but of different familiar size, the apparent distance of the second will be different from that previously reported for the first. The converse effect can also be shown, i.e., two objects of different sizes can appear at the same distance if their familiar sizes are appropriate. This experiment is also reported by Ittelson in the same paper referred to above.

Since the Hochbergs deal at some length with this paper by Ittelson, although only in general terms while considering it jointly with an earlier experiment by Hastorf (2), it may be worthwhile to examine the criticisms they make of this evidence. They write that "one can question first whether the effect was truly perceptual" (3, p. 109) and imply a negative answer to this question on the basis of evidence derived entirely from Hastorf's paper, although by implication it is extended to Ittelson's as well. However, they offer no evidence, and it can unequivocally be stated that there is no evidence to support this claim with respect to Ittelson's experiment. The second objection offered by the Hochbergs applies, by their own statement, only to Hastorf's experiment and therefore is not pertinent to the present discussion. It seems clear that the evidence offered by Ittelson, and by implication that of Ames on which it is based, is found "guilty by association" with Hastorf's experiment[2] and not by any concrete argument whatsoever. We may safely conclude that the Hochbergs have failed to offer sufficient reasons for discounting the evidence published by Ittelson (4) and by Ames (1) with respect to the cue of familiar size.

3. That the familiar size of a single object viewed monocularly in an otherwise dark room is an important determinant of its apparent distance can be shown by physiological measures. This fact is demonstrated in an experiment by Ittelson and Ames (7), and not referred to by the Hochbergs, in which both apparent distance and accommodation and convergence, as measured haploscopically, shifted in the direction predicted by the familiar size cue.

4. Two sets of three objects each can be constructed of the same physical dimensions so that the two sets produce identical retinal patterns. The apparent sizes and distances of the three objects in one set will nevertheless be quite different from those of the other set, if the familiar sizes of the objects in one set are different from those of the other set. This effect is shown in the "watch-card-magazine" demonstration of Ames which is described by Ittelson (6) and not referred to by the Hochbergs.

5. The size-distance perceptions related to a given stimulus can be changed by immediately prior experiences which change the familiar size attributed to that stimulus. This effect is illustrated in the "radial motion" demonstration of Ames which is described by Kilpatrick and Ittelson (8) and not referred to by the Hochbergs.

The solution of the rôle played by size in space perception, which is offered by the Hochbergs, provides such a simple and convenient resolution of the problems in this area that it is with regret that the author must report his inability to apply it to the above cases. If the Hochbergs, or any of their distinguished colleagues, can do so, we shall all be indebted to them.

Our failure in the above cases, however, should not be allowed to detract from the fact that the Hochbergs, by emphasizing the necessity of separating the effects attributable to relative size from those due to familiar size, have provided an important help in evaluating experiments involving the size cue.

The great value of applying this distinction can perhaps best be illustrated by referring to the experiment reported by the Hochbergs. "On two-dimensional 'reversible-screen' drawings . . . were drawn a man, 4¼ in. high, on one panel, and a boy of the same size and approximate contour, on the other . . . the problem is whether the panel with the boy appears nearer more of the time than that with the man, as might be expected from *familiar size*" (3, p. 110). The panel with the boy did not appear nearer more of the time, from which it is concluded that "an experimental separation of the two 'cues' found the latter [relative size] effective and the former [familiar size] ineffective" (3, p. 113).

---

[2] These comments are not in any way intended to imply that the criticisms of Hastorf's experiment are necessarily correct or justified but simply to point out that, correct or not, they simply are not relevant to the experiments reported by Ittelson and by Ames.

This conclusion, however, proves inadmissable, if we apply to it the Hochbergs' own criterion (unless they assume that relative size and familiar size are somehow separated when sizes are equated). Thus, if the Hochbergs take their own strictures seriously, it would seem that they should have written the statement quoted above to read, "the problem is whether the panel with the boy appears nearer more of the time than that with the man, as might be expected from *familiar size* or whether each panel appears nearer the same amount of time, as might be expected from *relative size*." Their conclusion, quoted above, might more properly have read, "in an experiment placing familiar size in conflict with relative size, the latter was found to be the more important determinant." Since it shows that under certain conditions relative size may be more important than familiar size (a conclusion which certainly no one questions) this experiment is primarily relevant to the study of conflicting cues and has no place in a discussion of whether or not familiar size can be shown to play a rôle in space perception.

Actually even the conclusion suggested above does not seem fully to be justified by the experiment, since the familiar sizes involved were the sizes of line drawings, although the entire discussion tacitly ignores this point by constantly referring to the "boy" and the "man." It is at least reasonable to ask whether a drawing of a boy is necessarily smaller than a drawing of a man, and whether the concept of familiar size has any meaning at all in this context. Possibly the "familiar size" of a line drawing is independent of the specific subject of the drawing. One cannot help wondering if the same results would have been obtained had real figures been used. While a young boy and a mature man of the same height (4¼ inches in the experiment reported) may be difficult to find, the effect of using real objects could very easily be obtained, for example, with the small playing card and the match box of the same size described by Ittelson (4). It is possible that if this were done the effect of familiar size might dominate that of relative size.

In addition to the above more fundamental considerations, two methodological questions remain to be asked concerning the experiment. First, in treating the data, the measure used was the difference between the total time one panel appeared nearer and total time the other panel appeared nearer, although it is clear that a more meaningful measure would have been the relative amount of time each panel was nearer. If the between individual, or between group, variability of total viewing time is great, this distinction becomes crucial since the use of the actual time difference gives undue weight to a subject whose reversals occur slowly, and hence who takes a long total time, relative to a subject whose reversals occur rapidly and who takes a short total time. This consideration is particularly important since two different groups were compared, and no evidence is offered that they were comparable in this respect. One can reasonably ask what would be the effect on the results of this experiment of using as a measure the differences between percentages of time nearer rather than the differences between total times nearer.

The second methodological question is simply whether or not the subjects in the experiment actually identified the drawings of the man and the boy as such; for example, a subject might have interpreted the boy in his play suit as a working-man in overalls. It is by no means suggested that this necessarily, or even probably, was the case in this experiment, but this is the kind of information which one would like to have in order properly to evaluate an experiment from which such wide-spread theoretical implications are drawn.

C. SUMMARY

In a recent paper, Hochberg and Hochberg state that they do not like theories which invoke the concept of familiar size as a cue to depth perception, and argue that the evidence for the existence of this cue is questionable since most of it hopelessly confuses familiar size with relative size. They report an experiment in support of the latter argument and offer a theory of size-distance perception based on "an autochthonous tendency toward homogeneity."

Reference is made to several experiments which are not considered in the paper by the Hochbergs. These experiments either successfully separate familiar size from relative size, or cannot be accounted for by an autochthonous tendency toward homogeneity, or both. The experiment reported by the Hochbergs is then discussed, and their conclusions are found to be unjustified both by their own criterion and on methodological grounds.

REFERENCES

1. Ames, A., Jr. Some demonstrations concerned with the origin and nature of our sensations (what we experience): A laboratory manual (preliminary draft). Hanover, N. H.: Hanover Institute, 1946 (mimeographed).
2. Hastorf, A. H. The influence of suggestion on the relationship between stimulus size and perceived distance. *J. of Psychol.,* 1950, **29**, 195-217.
3. Hochberg, C. B., & Hochberg, J. Familiar size and the perception of depth. *J. of Psychol.,* 1952, **34**, 107-114.

4. ITTELSON, W. H. Size as a cue to distance: Static localization. *Amer. J. Psychol.*, 1951, **64**, 54-67.
5. ———. Size as a cue to distance: Radial motion. *Amer. J. Psychol.*, 1951, **64**, 188-202.
6. ———. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285-294.
7. ITTELSON, W. H., & AMES, A., JR. Accommodation, convergence, and their relation to apparent distance. *J. of Psychol.*, 1950, **30**, 43-62.
8. KILPATRICK, F. P., & ITTELSON, W. H. Three demonstrations involving the visual perception of movement. *J. Exp. Psychol.*, 1951, **42**, 394-402.
9. LAWRENCE, M. Studies in Human Behavior. Princeton, N. J.: Princeton Univ Press, 1949.

Department of Psychology
Princeton University
Princeton, New Jersey

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 241-244.

CONTROLLING INTERACTION IN SMALL GROUP RESEARCH*

School of Public Relations and Communications, Boston University

MARTIN GROSSACK[1]

An experiment can be considered a contribution to group research without much hesitation when individuals interact in each other's presence, communicate verbally without restriction, and are able to form some impression of one another (1, 8, 11). These conditions certainly apply to the pioneer experiments in group settings (5, 13).

Recent developments in small group research methodology restrict the dynamic interaction of individuals and their forming of impressions of one another (7, 10, 12, 14). Yet such studies legitimately claim to treat processes that occur in small groups. Any research may be considered as a contribution to group psychology to the extent that the design permits interpretations at that level of abstraction.

This is, in certain respects, a discussion of the alleged artificiality of experimental research in social psychology. Experimental research has often been accused of lacking the "real-life" character of everyday behavior. The discussion of this allegation by Jahoda, Deutsch, and Cook (9) should receive the assent of both experimentalists and practitioners:

> Often the experimentation with emerging theoretical concepts and hypotheses is best performed in a laboratory, where, through careful isolating, some aspects of a problem can be studied better than when they blend into and are blurred by the full complexity of real-life situations.
>
> It will always be the case, on the other hand, that the relationships established in laboratory research of this nature will have been verified only for the limited range of conditions which it is possible to create in a laboratory setting. For virtually all problems of significance for the understanding of social relations, research must be conducted also, under the broader range of conditions found in the realistic settings of everyday life.

---

*Received in the Editorial Office on October 23, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This article is based on part of a Ph.D. dissertation submitted to the Graduate School of Boston University. The writer is indebted to Dr. Robert Chin for his raising of several issues basic to the problem and his helpful criticism, and to Dr. Anthony Stampolis, John W. Hill Professor, for supplying financial assistance from the Hill Fund in the preparation of this article.

TABLE 1
CONTROLS IN RECENT GROUP STUDIES

Author	*Topic*	*Controls*	*Source of data*
Crutchfield (2)	Assessment of persons through a quasi-group interaction technique.	Individuals cannot see one another. Communications supposedly by exchange of apparatus from one group member to another via experimenter serving as messenger. Actually, each individual receives standard stimuli.	Assessment of individuals on group squares test related to ratings and individual test performance.
Festinger, Schachter and Back* (13)	Social pressures in informal groups.	Field study of individuals.	Mainly an analysis of individual interview and sociometric responses in terms of residence units in a housing project.
Festinger and Thibaut (4)	Interpersonal communication in small groups.	Use of written communications without identity of sender.	Volume of communication; changes in card number representing opinion of subject.
Grossack (6)	Effect of cooperation and competition on cohesiveness, social influence and communication.	Individuals cannot see one another. Communications received were restricted to a standard set of stimulus messages. Communications "sent" were written messages during three "note-writing" periods.	Analysis of communications "sent." Sociometric and attitudinal questionnaire. Preferences on scale to human relations problem.
Heise and Miller (7)	Problem solving in small groups using various communication nets.	Restriction on content of verbal communications. Control of communication channel. Sender not allowed to indicate recipient.	Time necessary to complete task. Accuracy. Volume of communication. Construction of anagrams.
Kelley (10)	Communication in experimentally created hierarchies.	Communications received restricted to a standard set of stimulus messages. Communications "sent" were written messages.	Content analysis of communications "sent;" sociometric items group interview, errors in solving problem.
Leavitt (12)	Effect of communication patterns on group performance.	Control of communication channels. Use of written communications.	Time necessary to solve problem; number of errors, volume of communication, analysis of messages, questionnaire results.
Schachter, Ellertson, McBride and Gregory (14)	Cohesiveness and productivity.	Communications received were restricted to a standard set of stimulus messages. Communications "sent" were written messages.	Questionnaire results. Quantity of cardboards cut in group task.

*Though not an experimental study, this is included as an example of how knowledge concerning groups may be discovered without actually studying groups per se.

Recently developed methods of controlling group interaction present some of the most stimulating research techniques in contemporary social psychology. Table 1 describes controls in recent group research.

The adequacy of experimental research is dependent upon the extent to which independent variables are powerfully manipulated. This demands that other factors which could blur the operation of these variables be minimized.

The experiments described all controlled communication of the subjects in one way or another. This procedure gives the experimenter power to control his variables more rigorously than otherwise possible. Although none of these experiments studied groups in all their ramifications, all contributed to the scientific understanding of behavior in group settings.

At certain points in the scientific process, such sacrifices of scope are necessary. Group studies controlling interaction are worthwhile to the extent that they explore the possibilities of specific variables.

Specific techniques used have merits of their own. Planted communications successfully reduced restraining forces on the communication of irrelevant content for Kelley (10) and induced forces toward greater and less productivity (14) and structured interpersonal perceptions (6) in other studies. A new problem area is opened by investigating differences and similarities between note-writing and ordinary group situations. Heise and Miller (7) suggest several dimensions that may be studied by controlling communication channels.

One should not minimize the limited conditions of such laboratory research. Studies of groups without such controls, investigating variables developed experimentally, are equally necessary.

REFERENCES

1. BALES, R. F. Interaction Process Analysis. Cambridge: Addison-Wesley Press, 1950.
2. CRUTCHFIELD, R. S. Assessment of persons through a quasi group-interaction technique. *J. Abn. & Soc. Psychol.*, 1951, **46**, 577-588.
3. FESTINGER, L., SCHACHTER, S., & BACK, K. Social Pressures In Informal Groups. New York: Harper, 1950.
4. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. Abn. & Soc. Psychol.*, 1951, **46**, 92-99.
5. FRENCH, J. R. Organized and unorganized groups under fear and frustration. *Univ. Iowa Stud. Child Welf.*, 1944, **20**, 229-308.
6. GROSSACK, M. The Effect of Coöperation and Competition Upon Small Group Cohesiveness, Social Influence, and Communication. Ph.D. thesis, Boston University Graduate School, 1952.
7. HEISE, G., & MILLER, G. Problem solving by small groups using various communication nets. *J. Abn. & Soc. Psychol.*, 1951, **46**, 327-335.
8. HOMANS, G. D. The Human Group. New York: Harcourt Brace, 1950.

9. Jahoda, M., Deutsch, M., & Cook, S. W. Research Methods in Social Relations. New York: Dryden, 1951.
10. Kelley, H. H. Communication in experimentally created hierarchies. *Hum. Rel.,* 1951, **4**, 39-56.
11. Krech, D., & Crutchfield, R. S. Theory and Problems of Social Psychology. New York: McGraw-Hill, 1948.
12. Leavitt, H. Some effects of certain communication patterns on group performance. *J. Abn. & Soc. Psychol.,* 1951, **46**, 28-50.
13. Lewin, K., Lippitt, R., & White, R. Patterns of aggressive behavior in experimentally created "social climates." *J. Soc. Psychol.,* 1939, **10**, 271-299.
14. Schachter, S., Ellertson, N., McBride, D., & Gregory, D. An experimental study of cohesiveness and productivity. *Hum. Rel.,* 1951, **4**, 229-238.

89 Kingsdale Street
Dorchester 24, Massachusetts

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 245-252.

CUES, EXPECTATIONS, AND FIRST IMPRESSIONS*

School of Public Relations and Communications, Boston University

MARTIN GROSSACK[1]

A. INTRODUCTION

Social scientists have found it profitable to study the perceptions individuals have of one another. When the cognitive world of an individual is explored, the scientist is better able to predict and understand the person's future behavior. Examples of this are numerous both in the researches of scientists and the experiences of everyday living. Horowitz, Lyons, and Perlmutter (5), for example, have shown that the acceptance and rejection of an individual's communications in a meeting is related to how that person is perceived (his valence) by the other group members.

How people form impressions of one another has received relatively little attention in the scientific literature. Results that are available however, suggest that impressions are a function of both the perceiv*er* (his expectations) and the perceiv*ed* (initial cues) (1, 2, 4, and 6).

Deutsch (2), in discussing the interpersonal relations that develop in coöperative and competitive groups, predicted that "we should also expect the perceived source of these (cathected) activities to acquire, to some extent, a cathexis similar to that held with respect to actions." Experimentally, Deutsch demonstrated significantly more friendliness (as indicated by observers' ratings) within coöperative groups. His results support the point of view that perceptions of an individual are, to some extent, determined by initial cues.

Asch (1), and Haire and Grunes (4) found that, given a few adjectives (initial cues) about an individual, subjects are able to write personality sketches of them that are related to the specific constellation of the characteristics orginally induced.

Kelley (6) has clearly demonstrated that initial cues are used in formation

---

*Received in the Editorial Office on October 23, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This article is based on part of a Ph.D. dissertation submitted to the Boston University Graduate School. The writer would like to thank Drs. Robert Chin and John Thibaut for suggesting the use of stimulus messages in this research and their help in interpreting the data, Dr. Nathan Maccoby and Dr. Anthony Stampolis for many valuable suggestions, and the John W. Hill Fund of the Boston University School of Public Relations for financial aid in the preparation of this article.

of first impressions and has conceptualized the relation of expectations to impressions:

> A rôle expectation defines a set of specific goals for the observer and makes him dependent in certain ways upon other persons for reaching these goals. . . .
>
> Concern develops over characteristics which are relevant to aspects of the relationship. . . . Increased concern about certain characteristics will lead to the observer's looking for cues revelant to those characteristics (6).

The goal of this report is the presentation of some unexpected research findings (3) which support these earlier investigations.

B. Experimental Procedure

Eighteen student groups of five individuals each were created for experimental study. Individuals were assigned to a group only if they did not mention one another as friends.

Seats in the experimental room were so arranged that subjects upon entering and receiving places had to face a wall unable to see one another. Subjects were randomly assigned letters from *A* to *E*.

Each individual received a different set of mimeographed instructions, except for the fifth person, who was given one of the other four randomly. The types of instructions were: (*a*) high motivation, coöperative; (*b*) high motivation, competitive; (*c*) low motivation, cooperative; and (*d*) low motivation, competitive.

The experimental situation required the subjects to solve a problem in human relations and restricted their communications to written messages "delivered" for them by the experimenter. Actually, each subject received a set of identical notes supposedly written by the other four people in the group. The fifth person was necessary so that each individual could receive the same amount of notes (four).

The standard stimulus notes had the following thematic content:

Form A

> Shows high motivation to work on the group's problem, wants to work together with others in the group and to reach a common solution.
>
> *Sample message:* Johnny has had plenty of punishment—didn't get better. With Mr. O'Brien he was at his best behavior. All he needs is a more agreeable environment where he'll be treated sympathetically. How do you feel about it? I think we should all agree on one solution. This is fun.

Form B

Shows high motivation, wants to work on the problem alone.

Sample message: Well, I think working alone is the best procedure. A problem like Johnny's requires serious reflection. Please don't expect me to answer any notes you send.

Form C

Shows low motivation, is in field to the extent that communication centers about the problem and the sender's attitude toward it.

Sample message: Johnny tried to be a "good citizen" but he didn't know how to be or why he should be. Which solution we pick doesn't really matter to me.

Form D

Shows low motivation and is out of field, horseplays.

Sample message: Do you happen to know what's playing at the Loew's State this week?

C. Results

Sociometric items provided data concerning the consequences of the standard stimulus messages on judgments (rankings of contributions to the group meeting, friendliness, and preferences for future meetings) about the attributed "sender" of the notes.[2]

Cues were provided by the standard stimulus messages, while expectations were induced with the experimental instructions.

Responses to the sociometric items are presented in Tables 1 and 2. The items were:

> Please rank in order all the members of the group, including yourself, in terms of their contribution to the group's meeting.
>
> Which of the other members would you like to meet with again? Please rank.
>
> Please rank in order the other group members as to how friendly they were to you.

In Table 1 uniformities in response by subjects, regardless of the experimental instructions received, are presented for the three items.

"Senders" of high motivation, coöperative cues (attributed "sender" of Set *A* notes) were sociometrically ranked as contributing most to the group meeting. The difference in the rankings of high and low coöperative "individuals" (perceived as source of Set *C* notes) by all subjects was statistically significant at the .001 level of confidence. The "senders" of low motivation coöperative cues were ranked second, followed by the high and low motivation non-coöperative "senders" (perceived as source of Set *B* and Set *D* notes

---

[2] In the discussion that follows, "sender" refers to the attributed source of a message received by a subject.

TABLE 1
DIFFERENCES IN SOCIOMETRIC JUDGMENTS OF NOTE "SENDERS" IN TERMS OF CUES INDUCED

A. *Rankings of Contribution to Meeting*

CUES	A High Motivation Cooperative	C Low Motivation Cooperative	B High Motivation Competitive	D Low Motivation Competitive
1st Preference	72	2	4	0
2nd Preference	8	20	7	6
3rd Preference	3	34	16	10
4th Preference	2	20	29	30
5th Preference	0	8	30	40
Average Ranking	1.24	3.14	3.86	4.21

Statistical Analysis

A vs. B	t: 18.45	P: .001
A vs. C	t: 14.95	P: .001
A vs. D	t: 24.75	P: .001
B vs. C	t: 4.41	P: .001
B vs. D	t: 2.20	P: .05
C vs. D	t: 7.23	P: .001

B. *Preferences for Future Meeting*

CUES	A High Motivation Cooperative	C Low Motivation Competitive	B High Motivation Competitive	D Low Motivation Out of Field
1st Preference	69	6	3	6
2nd Preference	9	33	25	19
3rd Preference	2	36	16	25
4th Preference	3	8	25	24
Average Ranking	1.26	2.55	2.91	2.91

Statistical Analysis

A vs. C	t: 11.4	P: .001
A vs. B	t: 12.3	P: .001
A vs. D	t: 12.9	P: .001
C vs. B	t: 2.57	P: .02
C vs. D	t: 2.59	P: .02
B vs. D	t: 0.0	

TABLE 1 (*Continued*)

C. *Rankings of Friendliness*

CUES	A High Motivation Cooperative	D Low Motivation Out of Field	C Low Motivation Cooperative	B High Motivation Competitive
Ranked First	45	28	6	5
Ranked Second	20	18	36	10
Ranked Third	13	24	36	10
Ranked Fourth	4	11	8	56
Average Rank	1.71	2.22	2.53	3.44
Statistical Analysis	A vs. D	t: 3.31	P: .01	
	A vs. C	t: 6.30	P: .001	
	A vs. B	t: 12.01	P: .001	
	D vs. C	t: 2.17	P: .05	
	D vs. B	t: 8.13	P: .001	
	C vs. B	t: 6.70	P: .001	

respectively) in that order. *Regardless of the experimental instructions received, subjects responded similarly in judging individuals on the basis of available cues.*

Similarly, individuals seen as the source of coöperative notes were preferred for future meetings regardless of the experimental instructions received by the subjects. The order of preferences were (Table 1): high motivation, coöperative (mean: 1.26), low motivation, coöperative (mean: 2.55) and high motivation, competitive and low motivation, out of field (both means: 2.91).

Rankings of friendliness indicate that the high motivation, coöperative "sender" was perceived most friendly (mean 1.71); the low motivation, out of field (mean 2.22) next; followed by low motivation, coöperative, (mean 2.53); and high motivation, competitive (mean 3.44).

Table 2 presents differences in sociometric judgments of attributed "senders" of messages by recipients of different instructions (expectations).

Individuals receiving high motivation, competitive instructions were more likely to indicate sociometric preferences toward the "sender" of messages containing a high motivation, competitive theme. Also, individuals receiving low motivation, competitive instructions were more likely to indicate preferences toward the "sender" of messages containing low motivation, out of field themes than were the recipients of other types of instructions. There were no significant differences in rankings of coöperative "senders" for this item by recipients of different experimental instructions.

TABLE 2

DIFFERENCES IN SOCIOMETRIC JUDGMENTS IN TERMS OF EXPECTATIONS*

A. Preferences Toward "Sender" of High Motivation Competitive Cues For Future Meeting

Expectation: *Group I:*	*Mean*	*Group II:*	*Mean*	*t*	*P*
High Motivation Competitive	2.59	Low Motivation Cooperative	3.29	2.09	.05*
High Motivation Competitive	2.59	High Motivation Cooperative	3.21	2.16	.05

B. Future Meeting Preferences Toward "Sender" of Low Motivation Out of Field Cues

Expectation: *Group I:*	*Mean*	*Group II:*	*Mean*	*t*	*P*
Low Motivation Competitive	2.53	All Others	3.26	2.88	.01
Low Motivation Competitive	2.53	High Motivation Competitive	3.39	2.45	.02
Low Motivation Competitive	2.53	Low Motivation Cooperative	3.39	2.82	.01

C. Friendliness Ranking of "Sender" of High Motivation Cooperative Cues

Expectation:					
Competitives	1.44	Cooperatives	1.95	2.67	.02

*Group on left side is more favorable toward "note-sender."

In the friendliness judgments, there were no significant differences in rankings of individuals by recipients of different instructions, except for the rankings of the high motivation, coöperative "sender." Here, competitives significantly rank the high motivation, coöperative "sender" as more friendly than the individuals in coöperative situations judged her to be ("*t*": 2.67, *P*: .02).

D. INTERPRETATION OF RESULTS

The subjects in the present study did not know one another, received standard stimulus messages which provided different cues concerning the other group members, and made sociometric judgments of them. We may state the following ad hoc hypothesis as an explanation of the results in Table 1: *Initial cues are a determinant of impressions.*

Results in Table 2 suggest another ad hoc hypothesis: *Expectations are a determinant of impressions.* Some significant differences in judgments were made by the recipients of different instructions. Recipients of high and low motivation, competitive instructions responded more favorably to the perceived source of notes corresponding in content to their own instructions in making preferences for future meetings.

Coöperatives ranked the perceived source of notes with the high motivation, coöperative theme significantly less friendly than competitives ranked it to be. This may be explained by the fact that Set *A* notes' helpfulness and friendliness probably exceeded the expectations of competitive subjects (who could legitimately expect no assistance from others) but not coöperative.

Generally, these findings can be interpreted in terms of the point of view that behavior is a function of both the individual (expectations) and his environment (behavior of others). More specifically, the results reported here complement findings of Horowitz, Lyons, and Perlmutter (5), who reported some evidence that the acceptance or rejection of an individual's acts is related to the valence of the person perceived as source of the behavior that is judged.

In the present study, the individuals in each group were unknown to one another before the meetings. Valence was determined by the content of the communications received (cues) and the expectations of the recipients. Results reported here indicate that the valence of a person may be a function of the valence of his acts. While these results indicate how individuals *form* impressions, Horowitz, Lyons, and Perlmutter studied the *operation* of impressions in subsequent interaction.

Here, individuals receiving high motivation competitive and low motivation competitive instructions were more likely to prefer meeting the perceived source of cues corresponding to their initial expectations as induced by the experimental instructions. Competitive subjects were significantly more favorable than coöperatives in the friendliness rankings of the high motivation coöperative "note-sender." It appears that the expectations of competitive individuals were exceeded by the behavior of the attributed sender of coöperative messages. The amount of help and friendliness a competitive individual could legitimately expect was negligible (the success of a competitive individual required the failure of the other group members). For coöperatives to succeed in the experimental situation, it was necessary that their group be rated higher than any other group tested. This situation, in all likelihood, led to expectations of help and friendliness from others, which were not exceeded by the notes received from the high motivation coöperative "sender."

E. SUMMARY

A group experiment, using standard stimulus messages as an experimental control, provided some unanticipated data concerning the formation of first impressions. Individuals in the experimental situations were required to make sociometric judgments of the other group members. Since the subjects were unable to see one another or communicate verbally, available cues were restricted to the content of the standard stimulus messages substituted by the writer for notes actually written. Differential expectations were induced with high and low motivation and coöperative and competitive instructions.

Responses to the sociometric items supported the development of two ad hoc hypotheses:

1. Cues are a determinant of impressions.
2. Expectations are a determinant of impressions.

REFERENCES

1. ASCH, S. E. Forming impressions of the personality. *J. Abn. & Soc. Psychol.,* 1946, **41**, 258-290.
2. DEUTSCH, M. A theory of cooperation and competition. *Hum. Rel.,* 1949, **2**, 129-152.
3. GROSSACK, M. The Effect of Cooperation and Competition Upon Small Group Cohesiveness, Social Influence, and Communication. Ph.D. thesis, Boston University Graduate School, 1952.
4. HAIRE, M., & GRUNES, W. F. Perceptual defenses: processes protecting an organized perception of another personality. *Hum. Rel.,* 1950, **3**, 403-412.
5. HOROWITZ, M., LYONS, J., & PERLMUTTER, H. Induction of forces in discussion groups. *Hum. Rel.,* 1951, **4**, 57-76.
6. KELLEY, H. H. First Impressions in Interpersonal Relations. Ph.D. thesis, Research Center for Group Dynamics, Massachusetts Institute of Technology, 1948.

89 Kingsdale Street
Dorchester 24, Massachusetts

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 253-257.

TIME-PERSPECTIVE AND FEELING-TONE: A STUDY IN THE PERCEPTION OF THE DAYS*

Department of Psychology, University of Connecticut

MAURICE L. FARBER

A. PROBLEM AND PROCEDURE

By the very fact that we engage in psychological research, we almost perforce make the assumption that all human behavior is lawfully determined, though the laws may be as yet undiscovered. Given our research traditions however, we are apt to overlook the fact that everyday human activities, subjected to the proper methodological controls, can be as fruitful a source of hypotheses and laws as the most artfully contrived laboratory experiment. The present research seeks its data in that everyday life which is so omnipresent as to fade into shadowy background among the figure-grounds of our phenomenological world.

Man lives in a time-dimension, which he divides into units, usually on the basis of astronomical regularities. In Western society we have a recurrent pattern—the week—within which work and leisure are more or less uniformly distributed. For most of us there is, roughly, the work-week and the week-end.

That is the *objective* situation: but what of our perceptions of it? Man is driven by needs and thus oriented toward goals; one might expect gratifications and frustrations which would lead to different feeling tones toward different parts of the week. One hypothesis might be that individual differences in experience are so randomized throughout the days that no uniform pattern of feeling-tone toward particular days can be found. An alternate hypothesis, the one underlying the present study, would be that the uniform pattern of work and leisure results in some uniformity of feeling-tone toward different segments of the week.

Further, there present themselves several possible orientations of the dynamics: the simplest and most plausible, perhaps, is that the feeling-tone is mainly determined by the *present activities,* e.g., one's feeling-tone about Friday is related to the pleasantness or unpleasantness of one's activities on Friday. A second possibility is that the feeling-tone is related to the

---

*Received in the Editorial Office on November 7, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

immediate past, that is to the *past-time perspective,* e.g., one has a pleasant afterglow on Sunday because of a gay party the night before. Finally, the main determinant of the feeling-tone may be the future outlook, or *future-time perspective,* e.g., Friday may be pleasantly toned because one anticipates a week-end of leisure.

There is reason to hypothesize that the *future-time perspective* is of great importance in such a situation. The present writer has shown, in two quite disparate situations, that the future-time perspective is a major determinant of feeling-tone. In one study (2), it was found that the suffering of prisoners in a penitentiary was related, neither to the time they had already served nor to any aspects of the past-time perspective, nor yet to their particular present position in the prison, but quite markedly to an uncertainty in the future-time perspective. In the other study (3) the willingness of college students to advocate an immediate "showdown" with Russia, even at the risk of all-out war, was significantly related to the unfavorableness of their personal future outlook.

With regard to the week, with its rather uniform work-leisure distribution, a unique opportunity is presented for exploring the various hypotheses from a theoretical viewpoint. To this end, a questionnaire was devised and administered to 80 college students. In it, the respondents were asked to rank, if they could, the days of the week in order of preference, as well as to indicate in detail how they felt about each day of the week with their reasons. As an experimental variable, 28 of the respondents selected had university classes on Saturday, while the remainder did not.

B. RESULTS

Table 1 presents the mean ranks (1 = most liked, 7 = least liked) of the days of the week for the entire group as well as broken down for those with and without Saturday classes.

It will be noted first that the hypothesis that preferences would be random, depending upon individual experience, must be rejected. Regularities in preference certainly exist (only two subjects stated that they could not rank the days).

TABLE 1
MEAN PREFERENCE RANKS FOR DAYS OF THE WEEK

	Sun.	Mon.	Tues.	Wed.	Thurs.	Fri.	Sat.
All students	3.0	6.1	5.0	4.9	4.3	2.9	1.5
Students with Sat. classes	3.0	5.9	5.0	5.0	4.4	3.2	1.5
Students without Sat. classes	3.1	6.3	5.0	4.8	4.3	2.8	1.6

Contrary to any simple hypothesis that the leisure-days would be liked as a uniform block and the work-days similarly disliked, there are seen to be considerable differences among the work-days as well as between Saturday and Sunday. Actually, the work-day Friday is, for the entire group, preferred to the leisure-day Sunday! This suggests that factors other than the sheer activities on the days are involved.

For the entire sample, the order of preference is: Sat., Fri., Sun., Thurs., Wed., Tues., Mon.

That a valid order of preference is being measured is confirmed by the close agreement of the rank order found by Cason (1) and that found in the present study. Cason asked his students, to rate on an 11-point scale how they felt on the different days of the week. Despite this difference in technique, the Spearman rho between the two ranks is .89.

Saturday is the favorite day of the week. The reasons given for this choice stress, of course, that it is a free day, a day for dates and festivities. However, mention is also made of the fact that it is liked because it is *followed* by a free day, i.e., the *future outlook* on Saturday is favorable. Both factors contribute to the highly favored position of Saturday.

Monday is the least favorite day, confirming the folk-observation of "blue Monday." Since the actual work performed on Monday does not in general differ from that performed on the other week-days, the dislike of Monday cannot be attributed to the activities on that day. The explanation is again to be found in the future time perspective, as revealed in a typical comment (Subject 14): "On Monday the whole week of work stretches ahead of you."

From Monday through Friday a neat gradient exists with steadily rising outlook. The week-end might conceptually be viewed as a goal, with a goal-gradient of feeling-tone as it is approached.

The situation of Sunday is a fascinating one. Mention has been made in the psychoanalytic literature, by Ferenczi (4), for example, of "Sunday neurosis," based upon the observation that some people become inexplicably anxious and depressed on that day. The psychoanalytic explanation has been in terms of guilt associated with greater sexual temptations on Sunday, as well as the presence at home of the disciplinary father. In addition, for some people the work-week represents a reactive fleeing from forbidden instinctual fantasies, a solution impossible on Sunday. In the present study, Sunday is indeed found to have a curiously low rank, but at least part of the explanation would seem to be nearer at hand.

Sunday, we note, is rated not only considerably below Saturday, its week-end leisure-day associate, but is actually rated a little below Friday, a work-day! The reasons given for the Sunday ratings reveal the important rôle of the future-time perspective. Subject 22 states: "It (Sunday) brings to mind that I have an eight o'clock class next morning," and Subject 26: "It's dead and I think of Monday, which at that time seems terrible to me. However, when Monday does come, I enjoy it more than Sunday."

That Sunday represents the locus of conflicting forces is suggested by the great variation in rating it. The sigma for the mean Sunday rank is 1.9, as contrasted with 1.0 for Saturday, a difference which an F test reveals to be statistically highly significant. The pleasantness of a day of freedom is soured by the unpleasantness of the future outlook.

The operation of the various factors is seen more clearly when we compare those having Saturday classes with those having none. The ranks are given in Table 2.

TABLE 2

Students having Saturday classes	Students having no Saturday classes
Saturday	Saturday
Sunday	Friday
Friday	Sunday
Thursday	Thursday
Wednesday	Wednesday
Tuesday	Tuesday
Monday	Monday

Among those having Saturday classes Friday is rated lower than by those having no Saturday classes. This is congruent with the hypothesis of the importance of the future-time perspective. The outlook on Friday is not as rosy for those who have Saturday classes. The strong position of Saturday is not appreciably affected by the having of classes on that morning.

An individual analysis of deviant cases, i.e., those whose ratings differed markedly from the means, tends to confirm and define more sharply the dynamics which have been described. In one case, whether genuinely or as a neurotic symptom, the routine of the work-week is preferred to the week-end, and ratings are made accordingly. In several cases, there are unusual schedules, such as having a Saturday job, or having extraordinarily heavy or light programs on particular days. Almost without exception, it is possible to explain these cases by use of the same principles as with the larger group.

The present study, taken together with previous research would seem to suggest, with increasing confidence, that man's feeling-tone or mood is determined less by his present activities than by his future-time perspective. Western man, at least, seems indeed a creature governed by hopes and anxieties.

REFERENCES

1. CASON, H. General curves and conditions of feeling. *J. Appl. Psychol.*, 1931, **15**, 126-148.
2. FARBER, M. L. Suffering and Time-Perspective of the Prisoner. In *Authority and Frustration. Univ. Iowa Stud. Child Wel.*, 1944, **20**, 153-227.
3. ———. The Armageddon Complex: Dynamics of Opinion. *Pub. Opin. Quart.*, 1951, **15**, 217-224.
4. FERENCZI, S. Further Contributions to the Theory and Techniques of Psycho-analysis. London: Hogarth Press, 1926.

Department of Psychology
University of Connecticut
Storrs, Connecticut

Published as a separate and in *The Journal of Psychology*, 1953, 35, 259-266.

SIMPLE AUDITORY REACTION TIME IN AGED ADULTS*

Moosehaven Research Laboratory, Orange Park, Florida

WALTER D. OBRIST[1]

A. INTRODUCTION

Speed of performance on motor and intellectual tasks has generally been found to decrease in the later years of life (2). This fact is of considerable significance in the practical situation where the activities of the older person are evaluated or planned. It may also be of theoretical interest insofar as changes in speed are indicative of degenerative changes in the nervous system. The present experiment, part of a larger study on aging (3), is an attempt to determine trends during later life in the speed of performance on a simple motor task; namely, the classical finger reaction time test.

Simple digital reaction time to light or sound has been found by most observers to decrease with age until around 20 years, when it has its lowest value (6). Bellis (1) observed a subsequent increase in reaction time from middle age to 60 years. In what is perhaps the only study involving aged people, Miles (4) found little change before 60 years, but a substantial increase in the next two decades. Inspection of tables presented by both authors suggests a trend of greater individual differences with advancing age. Miles (5) places particular emphasis on differences between individuals, noting that about a quarter of the people in their seventies and eighties have, in spite of their age, reaction times that are faster than the general average for adults.

The present experiment was designed to check the finding of Miles that reaction time increases with age in later life, and to determine whether there is any increase in individual differences during this period. Whereas Miles employed only 15 cases over age 70, an attempt was made here to obtain greater statistical reliability through the use of a larger number of cases.

The possibility that brain deterioration in old age might be accompanied by disturbances in attention suggested an additional hypothesis; namely, that the performance of older people would be more variable. Thus, one

---

*Accepted for publication by Karl S. Lashley of the Editorial Board, received in the Editorial Office on November 9, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author is indebted to Dr. Robert W. Kleemeier for his suggestions concerning the analysis and interpretation of results.

would expect not only large differences *between* individuals, but also considerable variation *within* the individual when given a series of reactions to perform. This variation within an individual will be referred to here as *reaction time variability.*

B. Subjects

The reaction time test was given to a total of 141 male volunteers. Of these, 116 were residents of Moosehaven, a home operated by the Loyal Order of Moose for its aged members. Their ages ranged from 65 to 85 years. Only coöperative subjects in good health and relatively free from interfering sensory or motor defects were used in this sample. The remaining 25 volunteers were all members of Moose Lodge No. 455 in Jacksonville, Florida. Their ages ranged from 18 to 39 years. It is believed that the older subjects were initially comparable to the younger group in terms of intellectual ability and socioeconomic status.

C. Procedure

An auditory stimulus, consisting of a loud click which blended with the sound of a buzzer, was presented to the subject. The purpose of the click was to give the stimulus an instantaneous onset. By making the sound sufficiently loud, differences in auditory acuity among subjects were minimized. The response consisted of releasing the index finger of the dominant hand from a telegraph key. Releasing rather than pressing the key has been the procedure adopted by many investigators because of the accuracy in reproducing this movement in different experiments. Variations due to differences in size and spring tension of keys are thus minimized.

A "ready" signal was presented before each stimulus in order to give the subject the proper mental set. The time between this signal and the stimulus (the foreperiod) was varied randomly from one to two seconds, an interval found by other experimenters to yield the shortest reaction times. Two chronoscopes were used, one to record the length of the foreperoid in tenths of a second, and the other, the reaction time to the nearest five milliseconds. Figure 1 shows the circuit for this arrangement.

After receiving a standard set of instructions, the subject was given 25 practice trials. If there was any appreciable practice effect, as evidenced by continuing improvement, additional practice was given until the performance reached a constant level. This was then followed by 50 experimental trials, upon which the reaction time score of the individual was based. Frequent rest intervals were interposed in order to prevent fatigue.

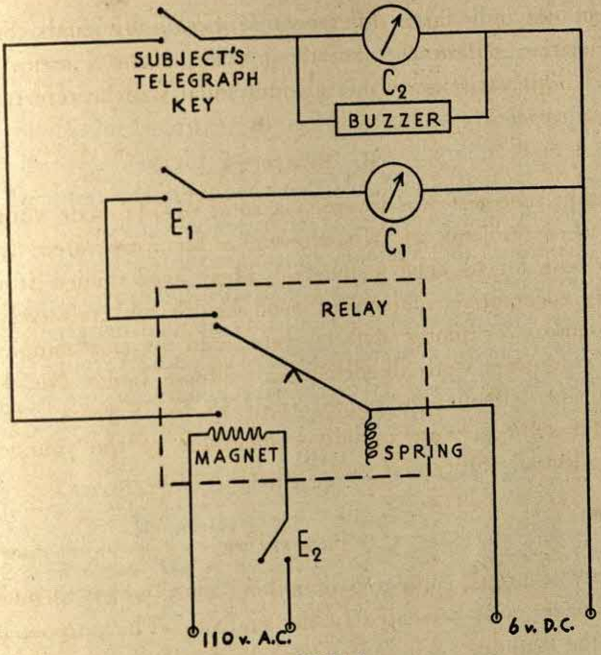


FIGURE 1

REACTION TIME CIRCUIT

Switch E_1 is closed by the experimenter at the moment the ready signal is given, thus activating the chronoscope, C_1. Observation of C_1 permits the experimenter to regulate the time interval beween ready signal and stimulus, which is varied from reaction to reaction. At just the right moment (as indicated by C_1) the stimulus is presented by closing E_2. Through the action of a relay, one circuit is opened, which stops C_1, and another circuit is closed, which starts C_2 and causes the buzzer to sound. The click of the relay blends with the sound of the buzzer, thus giving the stimulus an instantaneous onset. The subject, who has been pressing the telegraph key, reacts by releasing his hand when he hears the sound, and both the buzzer and C_2 are caused to stop. The reaction time and the length of the foreperiod are then obtained by reading the two chronoscopes.

Motivation was kept at a high level by praise and mild censure, and by encouraging the subject to make a game of it. Of necessity, individuals were handled in widely varied ways, depending upon their personality, motivation, and level of aspiration. Experience showed motivation to be one of the most crucial factors influencing reaction time, and so a particular effort was made to control it.

D. TREATMENT OF DATA

The 50 measurements made on each subject were plotted to form a frequency distribution. To facilitate the treatment of data, an individual's *reaction time* was defined as the median of this distribution of 50 measurements, and an individual's *reaction time variability* as the quartile deviation (semi-interquartile range) of this distribution. These statistics were used instead of the customary mean and standard deviation because of their greater stability, being influenced less by extreme measurements. Values lying beyond eight quartile deviations from the center of the distribution were omitted, since it was probable that they were the result of uncontrolled extraneous factors, and hence not true reaction times.

The reaction times and reaction time variabilities of the 141 individuals were then subjected to a statistical analysis. Comparisons were made between three age groups: Group I consisting of men between the ages of 18 and 39, Group II of men between 65 and 75, and Group III of men between 76 and 86. Table 1 gives the size and mean age of each of these groups.

TABLE 1
SIZE AND AGE OF THE THREE GROUPS TESTED

	Group I	Group II	Group III
No. cases	25	57	59
Age range	18-39	65-75	76-86
Mean age	27.5	71.5	80.6

E. RESULTS

Figures 2 and 3 show reaction time and reaction time variability plotted against chronological age. The three points determining each curve are the means for the three age groups. Note that both variables increase with age and that this increase is greater for the decade following age 70 than for the preceding 40 years.

Also shown in Figures 2 and 3 are the standard deviations for reaction time and reaction time variability (indicated by the dotted vertical lines). As can be seen, individual differences in both measures become greater with increasing age. Note particularly the proportion of cases in the two older groups whose reaction time or reaction time variability is less than the mean for the youngest group.

Table 2 summarizes the results, giving the exact values of points on the two curves. Statistical tests reveal that all differences between the means of adjacent age groups for both variables are significant at the .01 level of confidence, as shown by the *t*-values in this table. In the case of the standard deviation, note that for each variable one of the differences between age groups is significant at the .01 level.

A Pearsonian correlation coefficient of 0.71 was obtained between reac-

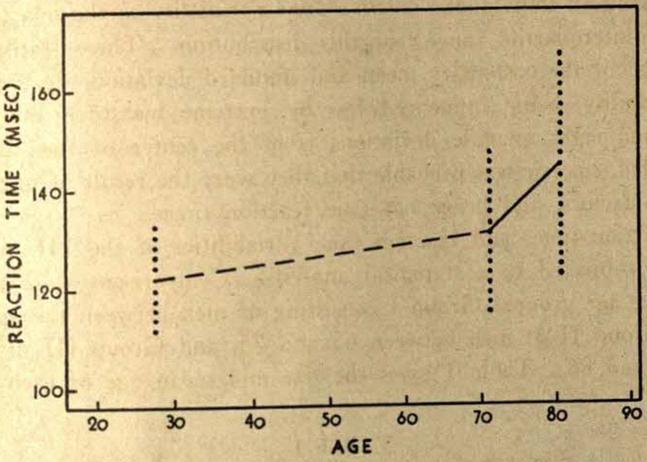


FIGURE 2
MEAN REACTION TIME VS. AGE
The mean reaction time for each of the three age groups is shown plotted against the average age of the respective group. The dotted verticle lines represent one standard deviation on each side of the mean.

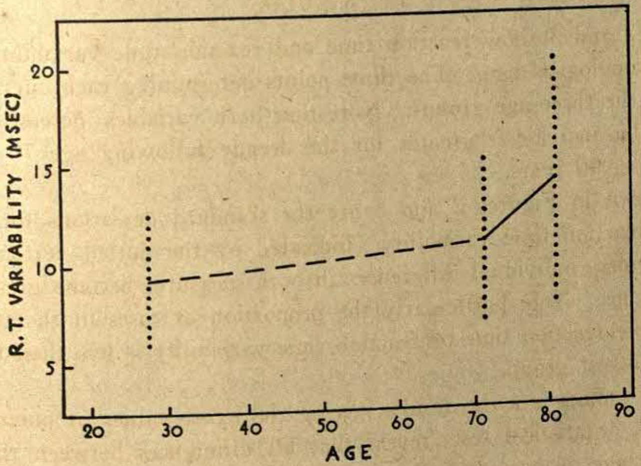


FIGURE 3
MEAN REACTION TIME VARIABILITY VS. AGE
Reaction time variability is a measure of variation in speed of reaction within the individual. The mean R.T. variability is the average individual variation for an age group, and is shown above plotted against the average age of the three different groups studied. The dotted verticle lines represent one standard deviation on each side of the mean.

TABLE 2

COMPARISON OF THE REACTION TIMES AND REACTION TIME VARIABILTIES OF THREE DIFFERENT AGE GROUPS

	Group I		Group II		Group III
Reaction Time †					
Mean	122.0		131.2		144.8
t		3.01**		3.72**	
SD	10.6		16.4		22.2
t		2.72**		2.24*	
Reaction Time Variability†					
Mean	9.3		11.0		14.0
t		6.26**		9.75**	
SD	3.3		4.1		6.2
t		1.34		3.08**	

t is the significance ratio for the difference between adjacent groups.

†Expressed in milliseconds.

*Significant at the .05 level of confidence.

**Significant at the .01 level of confidence.

tion time and reaction time variability when the three age groups were combined.

F. DISCUSSION

The present findings offer confirmation of Miles' results (4) on a statistically reliable basis. The reaction time of adults was found to increase with age, but the rate of this increase is apparently much greater during senescence than in the middle span of life. The trend toward larger individual differences in old age suggested by the Miles (4) and Bellis (1) data has also been confirmed.

An obvious explanation for the greater difference between individuals is that some people are declining rapidly, whereas others maintain their ability at a high level. This is certainly true on the physical level, where individual differences in rate of aging may be very pronounced. It is supported by the fact that in the present study as many as 32 and 15 per cent of the people in Groups II and III, respectively, had faster reaction times than the mean for Group I. In fact, the fastest performance in Group I was equalled by people in both Groups II and III. Age appears to be associated with little change in the performance of these individuals.

On the other hand, there were a sizable number of people in the two older groups (19 and 41 per cent, respectively) who had longer reaction times than the slowest individual in Group I. It is probably safe to assume that many of these latter people have undergone a considerable decline with age. A longitudinal study is needed to check this point and to show the exact nature of the decline in individual reaction time curves.

The significance of the greater reaction time variability in old age is difficult to assess. The procedures employed in the present experiment were intended to produce reaction times having a minimum variability and lying as close as possible to the physiological limit. Attempts were made to control variations in performance due to sensorimotor factors, practice effects, and fluctuations in motivation and fatigue. Because of the difficulty in controlling them, however, it is possible that some of these factors may have had a significant influence on reaction time variability.

Inspection of the individual distributions of reaction time measurements reveals a tendency for the older person to have a distribution that is skewed in the positive direction away from the physiological limit. For example, in Group III the average distance in milliseconds between the median and 90th percentile (on the slow reaction side of the distribution) is almost half again as large as the distance between the median and 10th percentile (on the fast reaction side). The young people in Group I, on the other hand, have more symmetrical distributions, the two distances being almost equal. The very slow reactions of old people that cause their distributions to be skewed usually occur randomly amongst average and faster reactions. This suggests the presence of some factor that interferes momentarily with the speed of performance.

It is the author's opinion that a central factor such as variation in attention is largely responsible for this momentary interference, accounting for both the skewness and increased variability of the old person's reaction time distribution. Obvious lapses of attention were numerous in some people, as inferred by frequent failures to respond to the stimulus. This is supported by the introspections of many of the subjects. Some reported that no matter how hard they tried they found it difficult to keep their mind consistently on the task. Disturbances in attention are part of the normal senile mental picture, and should probably be regarded as a primary factor influencing reaction time variability in this age group.

G. SUMMARY

Measurements of simple auditory reaction time and reaction time variability were obtained on a total of 141 male subjects at three different age levels. Analysis of group statistics revealed a significant increase in both measures with age, the rate of increase being greater after age 70. Individual differences in reaction time and reaction time variability also became greater with age, suggesting a differential rate of decline in the ability amongst old people.

REFERENCES

1. BELLIS, C. J. Reaction time and chronological age. *Proc. Soc. Exper. Biol. & Med.,* 1933, **30**, 801-803.
2. GRANICK, S. Studies in the psychology of senility—a survey. *J. Gerontol.,* 1950, **5**, 44-58.
3. KLEEMEIER, R. W. Third Annual Report of the Moosehaven Research Laboratory. Orange Park, Florida: Moosehaven, 1952. Pp. 80.
4. MILES, W. R. Correlation of reaction and coördination speed with age in adults. *Amer. J. Psychol.,* 1931, **43**, 377-391.
5. ———. Age and human ability. *Psychol. Rev.,* 1933, **40**, 99-123.
6. WOODWORTH, R. S. Experimental Psychology. New York: Holt, 1938. (Pp. 298-339.)

Moosehaven Research Laboratory
Orange Park, Florida

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 267-270.

ATYPICAL TONAL MEMORY*

College of Liberal Arts, Marquette University

CYRIL C. O'BRIEN

The interest of the author in the present title stemmed from witnessing an examination of a Baccalaureate student in music on his ability to hold in memory long sequences of sounds in increasing difficulty of successive, musical memory spans. The examination in question was one phase of a comprehensive, aural testing to which final, B.Mus. students of the music school of a particular university were subjected. The general procedure was initiated by the examiner, who played a number of notes on the piano—heard, but unseen by the student—after which the student wrote on music manuscript paper the correct pitch, rhythm, and time values of the notes sounded. The number of notes in each tonal pattern ran the gamut from as low as eight sequential sounds to as high as 20. The notes were written down by the student each time after one hearing. The sequences of notes played formed musical patterns involving the most frequently used musical intervals, although there were at least three or four tonal memory patterns, that comprised a number of diminished and augmented intervals. Such a test involved the immediate recall of unaccented, musical melodies. The test itself was difficult, particularly in view of the fact that individual differences in tonal memory span and musical memory span do exist even among music students especially gifted in memory for sound.

At this point an explanation of the title of this article is in order and a differentiation between tonal memory span and musical memory span as used by the author. In general, the article is concerned with ability in remembering sound—ability that deviates from the average. The article in the main will discuss and exemplify, by means of a brief survey, deviant, tonal memory behavior that is characterized by unusual facility in the retention and recall of lengthy, tonal patterns. As a simple example of a musical memory pattern, the remembering of "Doh, Me, Soh, Fah" after one hearing, may be cited. The ability to recall this and similar patterns of comparable, lesser, or greater degree of difficulty, after a single hearing, would constitute a four span musical memory. Musical memory spans are melodic in nature. A four-span tonal memory is ability to recall after a

---

*Received in the Editorial Office on November 9, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

single hearing any group of four tones that is a non-musical sequence. The line between a musical sequence and a non-musical sequence is not sharply drawn. In this study, the distinction is made by the author as an essential difference to indicate that tonal memory spans are designed sequentially in such a manner so as not to favor those who have had the advantages of musical training. This interpretation would indicate that musical memory spans, with their definite, melodic patterns, are obviously, other things being equal, easier for the musician than for the non-musician.

The items comprising both the Kwalwasser-Dykema test of tonal memory and the Drake Musical Memory Test have definite, melodic patterns. In the present study, however, the concept of tonal memory span is used.

In setting up a method of measuring memory for sound, Seashore (1) specifically used the memory span idea, and purposely designed his test items of various numbers of repeated tones to be generally unmusical in their interval order. One reason for the latter approach was to eliminate as far as possible the chances of rating a musical achievement rather than a musical talent. Of course, isolation of the factor of tonal memory and its evaluation under control highlighted the chief aspects of Seashore's experimentation.

Before presenting the sampling of those unusually gifted in memory for sound or tonal memory, and the method employed in this brief survey, a referral to some of the principal factors in Seashore's tonal memory tests—both Original and Revised versions—is made.

The Original issue of the tonal memory test comprised 50 items ranging from a comparison of two, two-tone patterns to a comparison of two, six-tone patterns. Each sequence of tonal patterns was played twice. In the second playing, one note only was changed. The listener identified the number of the tone that was altered in the second playing. The Revised Series of Measures of Musical Talents (1) contain a tonal memory test in Series *A* to be used along with the other tests of the battery for screening purposes, and a tonal memory test in Series *B*—for differentiating those of superior tonal memory from those of average or lesser talent. In the Series *B* tonal memory test, the most difficult list of items covers six-tone memory spans, and contains 30 possible responses.

From studies by Seashore, it is evident that the average tonal memory span is at least three tones. Those unable to hold two tones in memory are seriously deficient in this basic, musical trait. Ability of a six-tone memory span represents superior talent in this respect. The present status of research in this aspect of psychology suggests that tonal memory is a capacity rather than an ability, if not wholly, then fairly largely so. Since there is no question of the admixture of environmental influence from certain points of view, the word ability has been used to indicate tonal memory at times throughout this article.

After screening of 110 music students for superior ability in tonal memory, 22 were assumed to possess at least a six-tone memory span. This represents the upper limit of measurement for tonal memory in Series *B*—Seashore's Measures of Musical Talents. To this group of 22 was administered a specially designed test of 30 paired, non-melodic, tonal patterns. The term "non-melodic" must needs be interpreted with some provisos, since it is not always possible or agreed upon, what consists of melody or non-melody in a strictly musical sense. At least it can be stated that intervals were frequently employed in the diatonic and chromatic realms that ordinarily would not be met in the most commonly heard, musical melodies. The criteria for good, melodic sequence often fall in the areas of like, dislike, or preferential choice. The following sounds beginning with C(250-dv.) will serve as an example of a non-melodic pattern: C, F sharp, F natural, C sharp, E flat, G sharp, and B. When these notes are repeated with only number three altered—or any other single number only, changed—then it would indicate a seven-tone memory span, provided that there are a sufficient number of samples comparable in length and degree of difficulty.

The test of 30 trials designed by the author are in most instances of greater degree of difficulty than the items comprising the six-tone sequences of Seashore's tonal memory spans, since the express purpose of the writer was to discover cases of unusual tonal memory. This was attempted by means of a test including six paired 7-span patterns, six paired 8-span patterns, six paired 9-span patterns, six paired 10-span patterns, three paired 11-span patterns, and three paired 12-span patterns—all in all a total of 30 paired, non-melodic tonal patterns. These patterns were administered to 22 individuals at the rate of one tone per second on the piano, with a silent interval of three seconds between pairs.

Of the original 110 music students first screened in order to obtain those with the highest tonal memory span, 22 or 20 per cent were able to hold at least six tones (non-melodic) in memory. After administering the specially designed, 7 to 12 tonal span test to the select 22 from the original 110, the results in Table 1 were obtained.

From the music population sample studied, it is apparent that only 20 or about 18 per cent of the 110 students had better than a six-span tonal memory. Only one, or less than one per cent of the total number studied,

TABLE 1
NUMERICAL DATA OF ATYPICAL TONAL MEMORY OF 22 MUSIC STUDENTS

Tonal memory span		Number	Percentages
6		2	9.09
7		12	54.54
8		5	22.73
9		2	9.09
10		1	4.55
11		0	0
12		0	0
	Total	22	100.00

had a 10-tone memory span. None was able to hold in memory 11 or 12 non-melodic tones after a single hearing. Undoubtedly, there are individuals who possess even a greater tonal memory span than here indicated, but their numbers would probably be few.

It should be stressed that the testing for tonal memory span involved a single hearing only of the various, paired trials. In this respect a comparison might be made with testing of *digits forward*—as in the Stanford-Binet and Wechsler-Bellevue tests of intelligence—the number of digits one can recall after hearing the digits of a particular number recited once only at the rate of one digit per second. It is obvious that one or more repetitions of a number would help consolidate the sequence of the digits in the number. Similarly, repetitions of a tonal pattern favor overlearning in the extension of its memory span. However, the true, tonal memory span like the true, digit span implies immediate recall after each pertinent, auditory stimulus.

Since music students are a select group, statistically speaking, it is generally agreed that they have a greater tonal memory span than the average of the population at large. It would seem from the select group of 110, and the very select group of 22 further studied by means of a specially designed, non-melodic test to measure 7 to 12 tonal memory span ability, that phenomenal memory for lengthy, sound sequences of non-melodic patterns is a rare possession among the numerous other gifts with which the personality is endowed.

REFERENCE

1. SEASHORE, C. E. Manual of Instructions and Interpretations for the Seashore Measures of Musical Talents. Camden, N. J.: R.C.A.

2531 North Oakland Avenue
Milwaukee 11, Wisconsin

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 271-282.

METHODOLOGICAL CONSIDERATIONS UNDERLYING ELECTRODERMAL MEASUREMENT*

Department of Psychology, University of Southern California

WILLIAM W. GRINGS[1]

A. PROBLEM AND PROCEDURE

Many of the problems involved in the use of electrodermal (*GSR*) information in psychological studies arise from the methodological complexity of skin electricity measurements. Different methodologies exist in different laboratories, and each worker must collect such data with reference to his methods as are required by his particular research. There is no standard technique, and, with a few notable exceptions such as measurement units (18, 19, 21, 22), there is a relative scarcity of empirical information presented in support of methodological questions.

Historically, there is a great deal of information on the electrical nature of electrodermal phenomena (23, 24). Much of the original work, however, appeared in French and German periodicals, engineering journals, and physiological reports, all of which are somewhat inaccessible to psychologists. Concern has typically been for one of three electrical quantities. Of these, the most common is the resistance of the skin to the passage of a small direct current between two liquid or metal electrodes. A second is the impedance or reactance of the skin to an alternating current; and the third is the body-generated potential between the electrodes.

It is the purpose of this report to review some of the methodological considerations in the use of electrodermal data and to present some new material in relation to each. Most of these data were collected in the laboratory of the University of Southern California using undergraduate students as subjects.[2] The basic apparatus used was a DC amplifier, the characteristics of which are described elsewhere (33). Conventional input circuits were utilized, and outputs were registered variously by ink-writing milliammeter or cathode ray oscillograph. All circuits are believed to con-

---

*Received in the Editorial Office on November 11, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This study was aided by funds from the College of Letters, Arts, and Sciences and the Graduate School of the University of Southern California.

[2] The writer wishes to thank Dr. J. Stamm for helpful criticism, Dr. C. Lovell for assistance in securing subjects, and D. Clegg and B. Conklin for help in measuring subjects.

form to standard engineering practice and will be supplied on request but are not included here because of limitations on space.

Plastic cup electrodes were used with a zinc, zinc-sulphate, saline paste, skin junction. For potential measurement, the electrodes led directly into the amplifier input, and for other measurements they led into an appropriate input circuit. Three electrode placements were used: palm to palm (P-P), back to back (B-B), and palm to back (P-B), all on the right hand. Randomization of placement effects without moving the electrodes was made possible by the use of four electrodes and a selector switch.

B. Basic Skin Electricity

1. *Equivalent Resistance*

One important electrical question concerns the equivalent resistance circuit for the skin. By what combination of resistance and reactance components can the skin impedance situation be synthesized? The most widely accepted circuit is that of a combined series-parallel, capacity-resistance combination: a condenser in parallel with a resistance and both of these in series with other resistive components (23). The term "apparent" skin resistance is appropriate since all measurements include components attributable to the electrodes only.

Figure 1 shows the relation between *AC* base resistance and the frequency of measurement current for a sample of 10 subjects. This graph suggests the presence of an appreciable capacitive reactance component, which in turn indicates that the frequency of oscillation of the measurement current is an important parameter in the measurement picture. Most significant frequency effects occur below 1,000 cycles per second. The simplest equivalent impedance to this "average" subject is a .02 mfd. condenser in parallel with a 30,000-ohm resistance.

Comparable data to these were presented in 1933 by Brazier (2). In that case arm immersion rather than cup electrodes were used. Other observations appropriate to the interpretation of these data are the facts that different absolute results would be expected with, different electrode sizes (23) and that with reference to the effect of frequency upon magnitude of change on stimulation, as frequency increases impedance and capacity changes diminish (11, 27, 32).

2. *AC-DC Relation*

The practical question arises of what form of a relation exists between resistance to a direct current and the resistive component of impedance

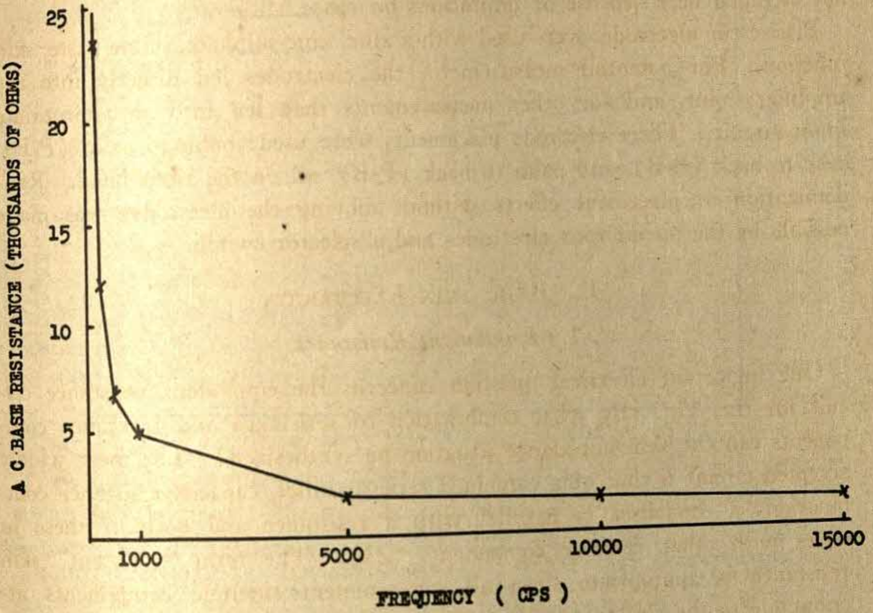


FIGURE 1
THE RELATION BETWEEN *AC* BASE RESISTANCE AND FREQUENCY OF MEASUREMENT CURRENT: EACH POINT REPRESENTS THE MEAN OF TEN SUBJECTS MEASURED FROM THE PALM TO THE BACK OF THE RIGHT HAND

to a 60-cycle alternating current. That is, if in one case a pure resistance *DC* bridge match is made and in the other case an impedance match is made using a single frequency of oscillation and balancing condensers, will the relation be simple and linear?

The measurements (palm-back) of *AC* and *DC* static states and measures of change for 32 subjects were plotted and tests of linearity made. In both cases the relation is clearly linear. The Pearson *r* between *AC* and *DC* static states is .92, whereas that for measures of change is lower ($r = .47$). Interpretation of the latter correlation is complicated by various factors such as the high relation which exists between static states and measures of change and the fact that changes in both capacity and resistance are involved in the values of change. The important fact here is that it appears reasonable to assume a linear relation in both cases. Since both *AC* and *DC* scales are currently in psychological use, this fact has bearing on the general problem of comparability of scales.

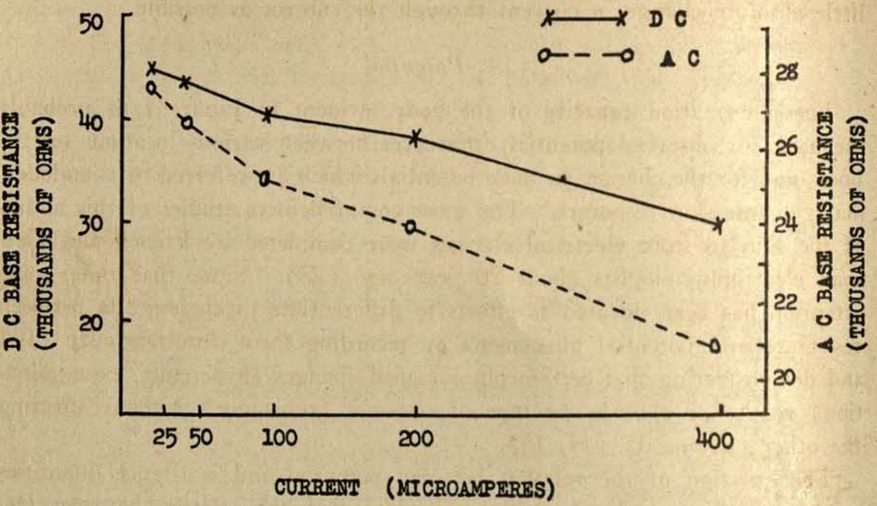


FIGURE 2

THE RELATION BETWEEN *AC* AND *DC* BASE RESISTANCES AND THE MAGNITUDE OF THE CURRENT PASSING BETWEEN THE ELECTRODES: EACH POINT REPRESENTS THE MEAN OF EIGHT SUBJECTS MEASURED FROM THE PALM TO THE BACK OF THE RIGHT HAND

3. *Current*

Another relationship which has significance in evaluating the comparability of results of different investigators is the relation between "apparent resistance" and the current passing through the body. It has long been known that measured skin resistance varies as the current passed through the skin varies, and the relation is usually assumed to be non-linear. There is, however, only one well known set of results on the matter (7), and this uses only three current settings.

To check on this relation, both the *AC* and *DC* base resistances of eight subjects were measured under five current levels, 25, 50, 100, 200, and 400 microamperes. The results for *AC* and *DC* measurements are shown in Figure 2. The *AC* measurements are 60 cycles, taken palm to back of the same hand. *DC* measurements are also palm-back.

These results confirm the need to control current as a variable in electrodermal measurements. Different investigators using identical apparatus but different measuring currents will not get comparable absolute measures. The problem is made even more important by the fact that resistance changes, or *GSR*s, are highly related to base values which in turn makes them vary as current varies. Because *GSR*s are typically inferred from current change devices, instruments for this purpose should operate with as little absolute change in current through the subject as possible.

4. *Potential*

The polarization capacity of the body, evident in Figure 1, is probably the basis for observed potential differences between various locations on the body, and for the changes in these potentials which are referred to as endodermal galvanic skin responses. The most comprehensive studies of this ability of the skin to store electrical charges were completed by French and German electrophysiologists about 20 years ago (23). Since that time, some attention has been devoted to efforts to differentiate psychologically between resistance and potential phenomena by recording them simultaneously (10) and demonstrating that certain physiological changes (hyperemia, exsanguination) reduce or abolish one type of response (resistance) without affecting the other (potential) (14, 15).

The question of the relation between potential and resistance quantities might next be raised. It will be recalled that the relation between *DC* measures and those of the resistive component of *AC* impedance were found to be linear and high. A similar plot of results with potential yielded quite different results. Eta correlations were first computed because of the non-linear nature of the plots, and Pearson *r*'s were later computed for comparison. The correlation (eta) between static *DC* and potential states was .59, between 60-cycle *AC* and potential .42, and between base capacity and base potential .40. Comparable values for measures of change were .47 between *DC* and potential and .39 between *AC* and potential. Pearson *r*'s on the same data were lower and negative. For all of these correlations the sample size was 32.

Potential changes on the skin following stimulation are particularly interesting because the electrical wave form representing the change is different from that of *DC* change. A brief period after the stimulus is administered there is typically a small potential change in one direction which is quickly followed by a major change in the opposite direction. Forbes (10) has studied this wave form and has named the first change an *"a"* response and the second a *"b"* response. The two changes apparently overlap, with a greater tendency for the *"b"* potential to be lessened by successive stimulation and to enter with strong stimuli. The *"a"* potential appears to be obtainable alone with both weak and strong stimuli. Forbes suggests that the two types of potential changes may represent different neural response mechanisms.

Persons who have worked with skin potentials have consistently emphasized the fact that body areas differ markedly in potential activity and that electrode-chemical and temperature variables are important determiners of the magnitude of potential and latency of change (4, 5, 8, 9). Some of this is evident in Table 1 where the distribution characteristics for base potential measures for the three-electrode placements are presented along with *AC* and *DC* values for the same subjects. Potential figures are in millivolts and are averaged without indication of polarity. It is of particular interest to note the difference in static potential from one electrode placement to another, with the palm to back position indicating the greatest difference. This does not agree with *AC* and *DC* resistance measures where in both cases the palm-back location has an intermediate value with the back-back difference greatest. Comparable data for measures of change are also presented.

C. Measurement Characteristics of Obtained Scores

1. *Frequency Distribution*

The summary information on static *AC, DC,* and potential values and measures of change presented in Table 1 indicates several features relevant

TABLE 1

Distribution Characteristics of Static or Base Values and Measures of Change: Changes Are the Average of the Responses to One Shock and One Tone Stimulus ($N = 33\text{-}34$)

Electrode Placement	Base measures P-P	B-B	P-B	Changes P-P	B-B	P-B
DC Resistance (Ohms)						
Mean	19,750	43,697	29,515	2,949	4,306	3,022
Median	17,500	33,500	26,300	2,100	2,980	2,820
SD	10,270	31,104	13,522	2,468	4,201	1,370
Q	5,500	6,830	5,500	1,220	1,778	937
AC Resistance (Ohms)						
Mean	13,095	22,000	19,032	1,230	905	1,067
Median	12,500	19,500	18,000	850	700	1,075
SD	5,132	8,077	6,013	1,331	904	735
Q	4,750	6,350	2,770	532	500	450
Potential (Millivolts)						
Mean	4.17	8.47	15.06	.927	1.003	2.094
Median	1.85	6.70	13.50	.525	.795	2.060
SD	4.67	8.40	9.52	.988	.955	1.408
Q	2.57	4.15	6.95	.355	.450	.995
Capacity (Microfarads)						
Mean	.016	.040	.032	Capacity changes were not measured.		
Median	.013	.032	.023			
SD	.009	.028	.022			
Q	.006	.009	.005			

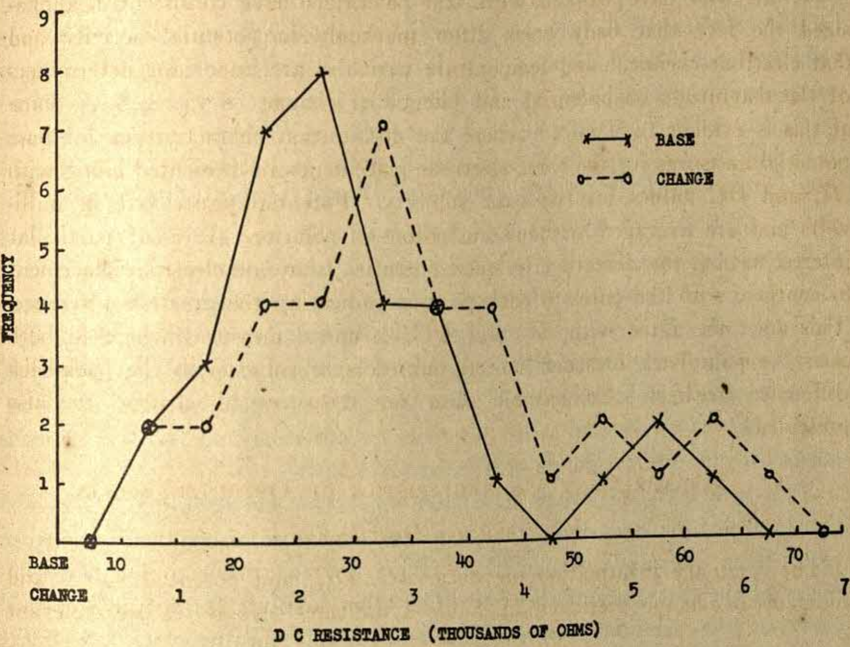


FIGURE 3
FREQUENCY DISTRIBUTIONS FOR *DC* BASE RESISTANCES AND MEASURES OF CHANGE TAKEN FROM THE PALM TO THE BACK OF THE RIGHT HAND

to the statistical manipulation of the data. The most obvious is the difference between positional and algebraic measures of central tendency and variability. In all cases the means exceed the medians in value, suggesting that the distributions are all positively skewed. The extent of this skew can be seen in Figure 3 where the unsmoothed frequency distributions for the *DC* measures are presented. The distribution curves for *AC* and potentials depart more from symmetry and normality than do those for *DC*. However, all distributions are unimodal, including potential measurements plotted either to retain or disregard their polarity (both positive and negative potential distributions pile up near the zero end of the scale).

The problem of non-normality of electrodermal data has been the subject of several published studies (19, 21, 22). Most of these have sought to cope with this distribution problem by transforming the scores into conductance or log conductance values. The evidence collected here suggests that the shape of the distribution is different for the different electrode placements (the palm-back placement giving a more symmetrical distribution than the other two). With measures of change the situation is complicated by the fact that the distribution shape is a function of stimulus intensity, for responses to mild stimuli pile up at the zero end of the scale.

Another fact worthy of note in the table is that the variabilities of the measures appear to change in the same fashion as the means and medians.

2. *Relation between Static Measures and GSRs*

A problem that is closely related to that of non-normality and is involved in proposed transformation of scores is the relation between the base or static values and measures of change. In general, the higher the base value the larger the *DC* resistance change. The correlation is often quite large. For these data, the eta coefficient computed between base and change values is .81 for the back-back electrode placement, .87 for palm-palm, and .48 for palm-back. The situation with reference to potential and *AC* quantities is not much different, although correlations are lower.

Two principal suggestions have been made for correcting for this relation by transforming scores. One is to use a measure of conductance change rather than resistance change, then taking the logarithm to increase symmetry of the distribution (19, 6). The other involves fitting the empirical curve of the relation between base and change measures with a linear function, then deriving a unit from the constants (18). Still another possibility exists in the use of the ratio of change to base (26).

3. *Derived Measures*

Several investigators use the electrical skin changes as a basis for deriving other *GSR* scales. Most of these endeavor to relate the different phases of the electrical *GSR* cycle into a ratio. Termed recovery measures, they are expressions of the magnitude of response as measured from base level to maximum divided by the amount of recovery (return to original level) which occurs after a stipulated time past the minimum has been reached (12, 13); or they represent the per cent resistance recovered in three seconds after the peak is reached (6). The angle in degrees between a tangent to the steepest part of the curve and the base line has also been used (6).

4. *Reliability of Electrodermal Data*

An important property of electrodermal data which has not been comprehensively studied is that of the reliability of the measurements. As applied to this laboratory situation the term reliability has several possible meanings. One involves how consistent the measures of the same individuals are under the same conditions on different days or different trials. This suggests two quite different estimates, one between days and the other between trials on the same day. The correlation between trials of the same day has been reported as high, numerous workers indicating values above .90 (20, 26, 31). Correlations between measures of change on shock and tone stimulation on the data here are .72 for *DC* measures of change, .84 for *AC,* and .65 for potential. Estimates for between days are not available, but it is reasonable to expect that they are much lower. It is also probable that correlations for static states exceed those for measures of change.

Since much of the criticism which has been leveled at the use of *GSR* in psychology has cited its "unreliability," a systematic study of the problem could do much in defining operations for controlling this highly variable quantity. Some of the difficulties inherent in its control are evident from the data presented here; others are associated with instrumentation variables; still others with complex subject variables such as set and instructions.

D. Conclusions

Within the confines of this study there appears to be no clear cut superiority of any of the three manifestations of electrical skin activity. *DC* measures show many advantages over *AC* and potential values. Their distribution form is more symmetrical and regular, and the experimenter need not worry about frequency effects, stray capacitance, and noise pickup (usually 60 cycles *AC*) by the subject. However, it is necessary with *DC* to be particularly attentive to electrode polarization problems. With capacitance factors balanced out, the base values of *AC* and *DC* are very highly related; but that relation drops for measures of change in which capacitance is permitted to vary. Both *AC* and *DC* are dependent upon measurement current. Potentials are somewhat harder to handle, since they involve base values which may differ in polarity from subject to subject and give typically a two-directional potential change upon stimulation.

All measures are subject to significant correlations between base values and those of change, and all vary greatly with different electrode placements. In addition, the distribution form of each measure varies with electrode placement. With the three placements used here, the palm-back distribution is most symmetrical and near-normal. Derived measures, while helpful in specific instances, do not offer a universal solution to these problems.

It should be noted that an important class of instrumentation variables was not systematically observed here. For example, "electrode resistance" is known to vary with time, current, concentration of electrolytes (the constancy of electrode moisture), the area of contact between the skin and the electrode, and the pressure of the electrode upon the skin (25, 17, 29, 9, 8, 1, 3, 5); and small differences in circuit design can greatly alter the variability of measures received (16). With all of these facts in mind, it is reasonable to conclude that the worker with electrodermal data is confronted with the choice of one of two alternatives. Either he must conduct a painstaking study of his proposed measurement method so that he can assure himself that the variance he is receiving is "subject" variance and not "error," or he must content himself with a study which assures him only of an absence of experimental bias and reduces him to the use of cruder tests of group or individual differences.

E. SUMMARY

It was the purpose of this study to compare three different manifestations of electrical skin activity with reference to their electrical properties, distribution characteristics, and interrelations. The three manifestations are *DC* resistance, *AC* impedance, and potential.

Samples of from 8 to 34 volunteer subjects were tested to supply quantitative estimates of the interrelations of electrical quantities in the measurement of electrodermal phenomena. Those for which data were presented include skin impedance as a function of frequency of measurement current, resistance versus reactance components of skin impedance, and the relation between skin resistance and the magnitude of the measuring current. The relation between *AC* and *DC* resistance components was shown to be a linear one, and the relation between skin potential and impedance quantities to be non-linear. Frequency distribution characteristics of static measures and measures of change were presented along with other statistical considerations.

REFERENCES

1. BLANK, I. H., & FINESINGER, J. E. Electrical resistance of the skin; effect of size of electrodes, exercise, and cutaneous hydration. *Arch. Neurol. & Psychiat.*, 1946, **56**, 544-557.
2. BRAZIER, M. A. B. A method for the investigation of the impedance of the human body to an alternating current. *J. Inst. Elect. Eng.*, 1933, **73**, 204-209.
3. BURNS, R. C. Study of skin impedance. *Electronics*, 1950, **23**, 190-196.
4. DARROW, C. W. Simultaneous galvanic reactions in different skin areas. *Psychol. Bull.*, 1932, **29**, 674-675.
5. ————. Considerations for evaluating the galvanic skin reflex. *Amer. J. Psychiat.*, 1933, **13**, 285-298.
6. ————. The equation of the galvanic skin reflex curve: I. The dynamics of reaction in relation to excitation background. *J. Gen. Psychol.*, 1937, **16**, 285-309.

7. Davis, R. C. Factors affecting the galvanic skin reflex. *Arch. Psychol.*, 1930, **18**, No. 115.
8. Floyd, W. F., & Keele, C. A. Further observations on changes of potential and EMF recorded from the human skin. *J. Physiol.*, 1936, **86**, 23-25P.
9. ———. Electrolytic phenomena in electrodes used for physiological purposes. *J. Physiol.*, 1936, **86**, 25-27P.
10. Forbes, T. W. Skin potential and impedance responses with recurring shock stimulation. *Amer. J. Physiol.*, 1936, **117**, 189-199.
11. Forbes, T. W., & Landis, C. The limiting *AC* frequency for the exhibition of the galvanic skin (psychogalvanic) response. *J. Gen. Psychol.*, 1935, **13**, 188-193.
12. Freeman, G. L. The Energetics of Human Behavior. Ithaca: Cornell Univ. Press, 1948. (Chapt. III.)
13. Freeman, G. L., & Katzoff, E. T. Methodological evaluation of the *GSR* with special reference to the formula for the Recovery Quotient. *J. Exp. Psychol.*, 1942, **31**, 239-248.
14. Goadby, K. W., & Goadby, H. E. The nervous pathway of the psychogalvanic reflex. *J. Physiol.*, 1949, **108**, 177-184.
15. ———. Simultaneous photographic records of the potential and resistance effects of the psycho-emotive response. *J. Physiol.*, 1936, **86**, 11-13P.
16. Greenwald, D. U. Circuits now available for the measurement of electrodermal responses. *Psychol. Bull.*, 1935, **32**, 779-791.
17. ———. Electrodes used in measuring electrodermal responses. *Amer. J. Psychol.*, 1936, **48**, 658-662.
18. Haggard, E. A. Experimental studies in affective processes: II. On the quantification and evaluation of measured changes in skin resistance. *J. Exp. Psychol.*, 1945, **35**, 46-56.
19. ———. On the application of analysis of variance to *GSR* data: I. The selection of an appropriate measure. *J. Exp. Psychol.*, 1949, **39**, 378-392.
20. Jeffress, L. A. Galvanic phenomena of the skin. *J. Exp. Psychol.*, 1928, **11**, 130-140.
21. Lacey, O. L. An analysis of the appropriate unit for use in the measurement of level of galvanic skin resistance. *J. Exp. Psychol.*, 1947, **37**, 449-457.
22. Lacey, O. L., & Siegel, P. S. An analysis of the unit of measurement of the galvanic skin response. *J. Exp. Psychol.*, 1949, **39**, 122-127.
23. Landis, C. Electrical phenomena of the skin. *Psychol. Bull.*, 1932, **29**, 693-752.
24. Landis, C., & De Wick, H. N. The electrical phenomena of the skin (*PGR*). *Psychol. Bull.*, 1929, **26**, 64-119.
25. La Porte, H. R. Electrode factors influencing the measurement of the galvanic skin response. Master's thesis, University of Southern California, 1951.
26. Lauer, A. R. Reliability of the galvanic reflex. *Amer. J. Psychol.*, 1929, **41**, 263-270.
27. McClendon, J. F., & Hemingway, A. Variations in the polarization capacity and resistance of the skin. *J. Gen. Physiol.*, 1930, **13**, 621-626.
28. ———. The psychogalvanic reflex as related to the polarization capacity of the skin. *Amer. J. Physiol.*, 1930, **94**, 77-83.
29. Rothschild, Lord. The polarization of a calomel electrode. *Proc. Roy. Acad. Sci., London*, 1938, **125**, 283-290.
30. Seward, J. P., & Seward, G. H. The relation of galvanic skin reactions to preceding resistance. *J. Exp. Psychol.*, 1935, **18**, 64-79.
31. Shock, N. W., & Coombs, C. H. The reliability and validity of some indices of changes in skin resistance as a measure of affectivity. *Psychol. Bull.*, 1937, **34**, 559-560.

32. THOULESS, R. H. Magnitude of the psychogalvanic reflex phenomenon in alternating current circuits and with direct current of high EMF. *Proc. IX Int. Cong. Psychol.,* 1929, **9**, 432-433.

33. TRUEBLOOD, H., & GRINGS, W. W. A *DC* amplifier for *GSR* and low frequency potential measurement. *J. of Psychol.,* 1950, **30**, 401-403.

34. WELLS, F. L., & FORBES, A. On certain electrical processes in the human body and their relation to emotional reactions. *Arch. of Psychol.,* 1911, No. 16, 1-39.

Department of Psychology
University of Southern California
3518 University Avenue
Los Angeles 7, California

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 283-298.

FACTORS INFLUENCING ROUTINE PERFORMANCE UNDER NOISE: I. THE INFLUENCE OF "SET"*

Institute of Educational Research, Indiana University

E. VICTOR MECH

A. THE PROBLEM

Industry and educational institutions have an interest in the behavior of individuals. They are interested primarily in providing conditions which are conducive to maximum output without undue effort or fatigue. The emphasis which is given to noise abatement campaigns as indicated by Ryan (11) implies a common assumption that noise is a hindrance to effective work. This interest in controlling noise has, however, far outrun the empirical evidence available.

Further justification of Ryan's criticism is to be found in an earlier review by Berrien (2) who indicates that in the light of this widespread interest, one might reasonably expect to find a considerable body of well grounded scientific literature relevant to the topic of noise and how it affects the behavior of man. He points out, however, that there is a lack of evidence dealing with the problems at the human level in terms which provide a basis for generalizations.

One of the reasons for the inconclusive nature of the data concerning the effects of noise seems to be in the very nature of the experimental situations in which a great deal of the data have been gathered. That the subject knows he is being tested for the effects of distraction or the strength of his ability to "concentrate," is an unfortunate aspect of the experiments in this area. Such stimuli as gongs, auto horns, electric shock, loud noises, and acrobats performing in the same room with the subject should not be expected to render adequate data on the effects of distraction on behavior. Indeed, these operations may, instead, be an indication of momentary disorganization of behavior upon the accomplishment of a specified task.

We feel with Berrien (2), that how important the variable, "suggestion" could be in such situations, is not clear. Baker (1) has reported positive results with male subjects, who worked for 5-minute periods, under music and quiet conditions. Baker concludes that output can be lowered or raised

---

*Received in the Editorial Office on November 15, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

dependent upon what he terms pre-task "sets." Berrien is critical of this finding on the basis that since neither the work nor the length of the work spells in Baker's study are comparable to the industrial situation there is no reason to expect that "suggestion" would be a major factor influencing personal efficiency.

Berrien seems to have misinterpreted Baker's results somewhat, which were not concerned with personal "efficiency," but with overt performance changes in a specified task. It is difficult to see how Berrien can reconcile efficiency with performance. Because one's performance rate is raised or lowered is not sufficient evidence for postulating that we have here an increase or decrease in "efficiency."

If we do not know anything about the cost of work to the individual we are on more tenable terrain if the data are reported in terms of performance changes, as Baker has done, and avoid using the term "efficiency."

Berrien's critique of "suggestion" as an important variable in work tasks, seems justified, on the other hand, that the work periods used were rather limited, and not comparable to industrial situations.

It is doubtful, though, whether Baker designed his experiment with an "industrial situation" in mind.

Most of the experiments dealing with noise adaptation, have used physiological measures. These experiments (4, 5, 7, 8, 9, 10, 13) seem to suggest that there is little effect of noise either upon output or upon effort expenditure during work. This appears to be at least one interpretation of the data. Actually, the factor of "set" has received meager experimental treatment with regard to being a factor in influencing overt behavior under noise conditions. In addition, very little is known of the effects of the more "meaningful" noises, such as verbal conversation or of radio programs on subsequent performance.

The experiment being reported deals with two facets: (*a*) whether systematic performance changes are possible, in the presence of operational noise stimuli, as a function of "set"; (*b*) to gain an indication of the relative permanence of a "set" under specified conditions.

B. PROCEDURE

The purpose was to design a situation which would yield some precise data with respect to the information desired. Four groups of 15 *S*s each were required to perform a verbal addition task for 30 minutes a day.

Each group worked eight days under two conditions each day; one a noise condition, the other a quiet condition. Each condition on each day consisted of 15 60-second work periods. Thus each individual *S* served a total of four work hours in the experiment, over the eight-day period. The task employed in the investigation has been used by Starch (12) and other experimenters dealing with human learning. The task consisted of adding 6, 7, 8, and 9 successively to a given two-place number and then repeating this operation for 60 seconds. The addition would then be stopped by the experimenter and *S* given another two-place number. Each of these 60 seconds of addition will be termed a "problem." All *S*s were told to add as rapidly and accurately as possible; and that credit would be given for the total number of correct responses. The number of correct responses during each 60-second period will be designated as the basic response measure. Clearly then, there were 15 such measures under noise and 15 measures under quiet for each *S*, each day.

The verbal noise used in the experiment was Volume I: "I Can Hear It Now," side No. 1. This record contains verbal narrations of important news events between the 1933-1945 era. The recording was played at an average intensity of 70 decibels, as measured by a sound level meter, Model 759-B. The measurement of 70 db. was made at the position the *S* was sitting. The 60 *S*s used in the experiment were from an undergraduate section in Educational Psychology. Participation in the experiment was required.

Each *S* chose the session which was most appropriate to his particular schedule, this was of paramount importance because the experimental conditions required attendance the same time each day for eight sessions. There was a 24-hr. interval between all replications, except that between 5 and 6. The time interval between the fifth and sixth replication was 72 hrs. Noise and quiet was alternated in a simple *ABBA* order. The experimental conditions for the four groups were as follows:

Group A. The *S*s in this group were told only that the experiment concerned the effects of noise on work. They were told nothing in relation to whether noise should facilitate or inhibit their performance.

Group B. *S*s in this group were told that the experiment concerned the effects of noise on work. They were shown a faked work graph, which was passed off as genuine, and that the results were from a previous experiment in this area. These curves indicated that in the previous experiment other *S*s had performed better under noise conditions, than under quiet.

Group C. These *S*s were treated the same as those in Group *B*, except that the faked work curves demonstrated that previous *S*s had performed better under quiet conditions.

Group D. The *S*s in this group were treated the same as those in Group *C*, except that the spurious work curves demonstrated that previous *S*s had at first performed better under quiet conditions for the first three or four days and then gradually getting accustomed to the noise, performed better under the noise conditions for the last three or four days.

The "Sets" were not introduced until prior to the second session, thus an initial measure was obtained on Day 1, for all *S*s in order to utilize the covariance technique. No knowledge of results was given to any *S* at any time.

The systematic variation of the experimental instructions was made according to the schema in Table *A* in which *A, B, C,* and *D* represent the different "treatments."

TABLE *A*
SCHEMA: SYSTEMATIC VARIATION OF INSTRUCTIONS

Set I	Set II	Set III	Set IV
A	B	C	D
B	D	A	C
C	A	D	B
D	C	B	A

C. RESULTS

Table 1 presents the composite performance measures of the four groups under noise and quiet conditions, while Figures 1 and 2 show the plot of these measures. Tables 2 and 3 contain the covariance *F* values under noise and quiet for Days 2, 5, 6, 8. It is clear from the graphs that performance differences existed between the groups, and that several inversions have taken place. The "Sets" were not introduced until Day 2 and each covariance computation was made against the initial measure, Day 1.

The next operation was to consider the intra-group differences with re-

TABLE 1
THE COMPOSITE PERFORMANCE SCORES OF FOUR GROUPS OF SUBJECTS UNDER NOISE AND QUIET CONDITIONS, FOR EIGHT REPLICATIONS

"Set"	Conditions	Replications							
		1	2	3	4	5	6	7	8
A	Quiet	2,581	3,001	3,539	3,959	4,189	4,635	4,697	4,209
	Noise	2,294	3,029	3,364	4,050	4,092	4,504	4,529	4,160
B	Quiet	2,806	2,700	2,988	3,451	3,302	3,830	3,643	4,207
	Noise	2,590	3,743	3,664	4,301	4,587	5,257	4,876	6,372
C	Quiet	2,371	3,348	3,961	4,123	5,287	4,643	5,156	4,968
	Noise	2,976	2,702	3,156	3,300	3,844	3,782	4,076	3,934
D	Quiet	2,612	3,122	3,312	3,681	3,197	4,140	4,083	4,208
	Noise	2,514	2,425	2,730	3,337	3,682	4,965	4,602	5,304

TABLE 2
Covariance *F* Values Under Noise for the Four Groups of *S*s

Days	1-2	1-5	1-6	1-8
F	7.87**	3.81**	1.26	9.96**

**Significant beyond the .01 level of confidence.

TABLE 3
Covariance *F* Values Under Quiet for the Four Groups of *S*s

Days	1-2	1-5	1-6	1-8
F	2.68	6.47**	4.30**	1.13

**Significant beyond the .01 level of confidence.

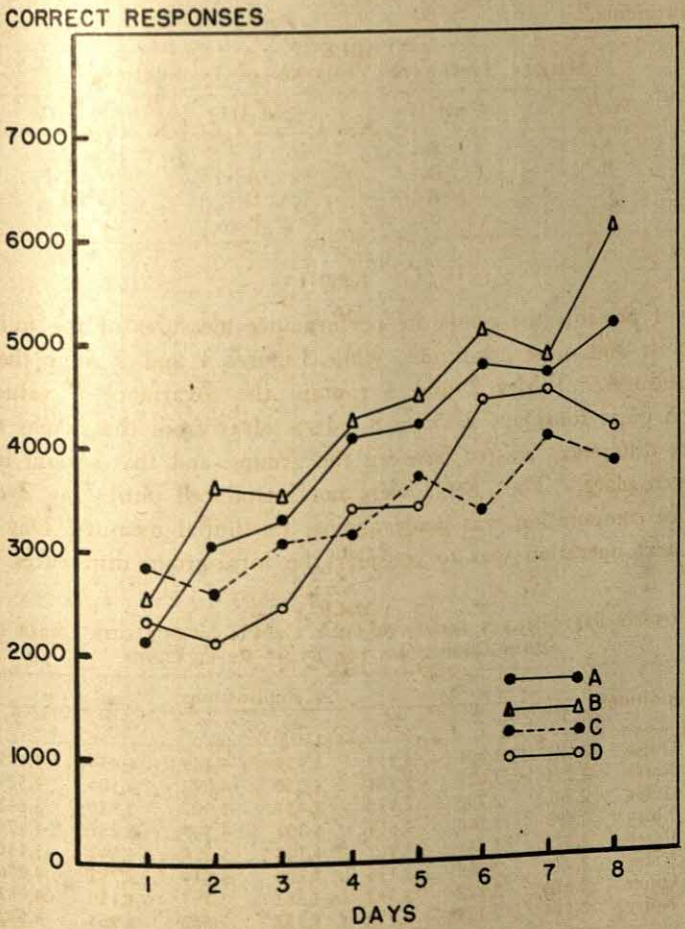


FIGURE 1
Performance Curves of Four Groups of *S*s Under Noise for An Eight-Day Period

TABLE 4

CROSS-OVER ANALYSIS OF VARIANCE FOR COMPARING THE PERFORMANCE OF GROUP *A* *S*s UNDER NOISE AND QUIET

Rows	Replications 1	2	3	4	5	6	7	8	Totals
First	Q172.06	N201.93	Q235.93	N270.00	Q279.26	N300.27	Q313.13	N277.33	2,049.91
Second	N152.93	Q200.07	N224.27	Q263.93	N272.80	Q309.00	N301.93	Q280.60	2,005.73
Totals	324.99	402.00	460.20	533.93	552.06	609.27	615.06	557.93	4,055.64

Treatment totals N2,001.3; Q2,054.1

Analysis of Variance

Source of variation	*df*	Sum of squares	Mean square
Columns (replications)	7	36,221.68	5,251.59
Rows (first versus second)	1	123.21	123.21
Treatments	1	174.24	174.24
Error	6	560.05	93.34
Total	15	37,079.18	

F-ratio for Treatments = 1.86; not significant.

TABLE 5

CROSS-OVER DESIGN FOR COMPARING THE PERFORMANCE OF GROUP *B* *S*s UNDER NOISE AND QUIET

Rows	Replications 1	2	3	4	5	6	7	8	Totals
First	Q187.06	N249.53	Q199.20	N286.73	Q220.13	N350.46	Q246.20	N424.80	2,164.11
Second	N172.66	Q180.00	N244.26	Q230.06	N305.80	Q255.33	N325.06	Q280.46	1,993.63
Totals	359.72	429.53	443.46	516.79	525.93	605.79	571.26	705.26	4,157.74

Treatment totals: N2,359.30; Q1,798.44

Analysis of Variance

Source of variation	*df*	Sum of squares	Mean square
Replications	7	42,043.20	6,006.17
Rows (first versus second)	1	1,806.00	1,806.00
Treatments	1	19,600.00	19,600.00
Error	6	5,456.83	909.48
Total	15	68,906.03	

F-ratio for treatments = 21.55; significant at .01 level.

TABLE 6

CROSS-OVER DESIGN FOR COMPARING THE PERFORMANCE OF GROUP *C Ss* UNDER NOISE AND QUIET

	Replications								
Rows	1	2	3	4	5	6	7	8	Totals
First	Q158.06	N180.13	Q264.06	N220.00	Q352.46	N252.13	Q343.73	N262.26	2,032.83
Second	N198.40	Q223.20	N210.40	Q274.86	N256.26	Q309.53	N271.73	Q331.20	2,075.58
Totals	356.46	403.33	474.46	494.86	608.72	561.66	615.46	593.46	4,108.41

Treatment totals: N1,851.11; Q2,257.10

Analysis of Variance

Source of variation	*df*	Sum of squares	Mean square
Replications	7	33,438.10	4,776.87
Rows (first versus second)	1	114.10	114.10
Treatments	1	10,302.25	10,302.25
Error	6	5,597.29	932.88
Total	15	49,451.74	

F-ratio for treatments = 11.04; significant at .01 level.

TABLE 7

CROSS-OVER DESIGN FOR COMPARING THE PERFORMANCE OF GROUP *D Ss* UNDER NOISE AND QUIET

	Replications								
Rows	1	2	3	4	5	6	7	8	Totals
First	Q174.13	N161.66	Q220.80	N222.46	Q213.13	N221.00	Q272.20	N207.46	1,692.84
Second	N167.60	Q208.13	N182.00	Q245.40	N180.80	Q276.00	N237.60	Q280.53	1,778.06
Totals	341.73	369.79	402.80	467.86	393.93	497.00	509.80	487.99	3,470.90

Treatment totals: N1,580.12; Q1,890.32

Analysis of Variance

Source of variation	*df*	Sum of squares	Mean square
Replications	7	14,368.54	2,052.65
Rows (first versus second)	1	453.70	453.70
Treatments	1	6,014.00	6,014.00
Error	6	1,067.64	177.96
Total	15	21,903.88	

F-ratio for treatments = 33.70; significant at .01 level.

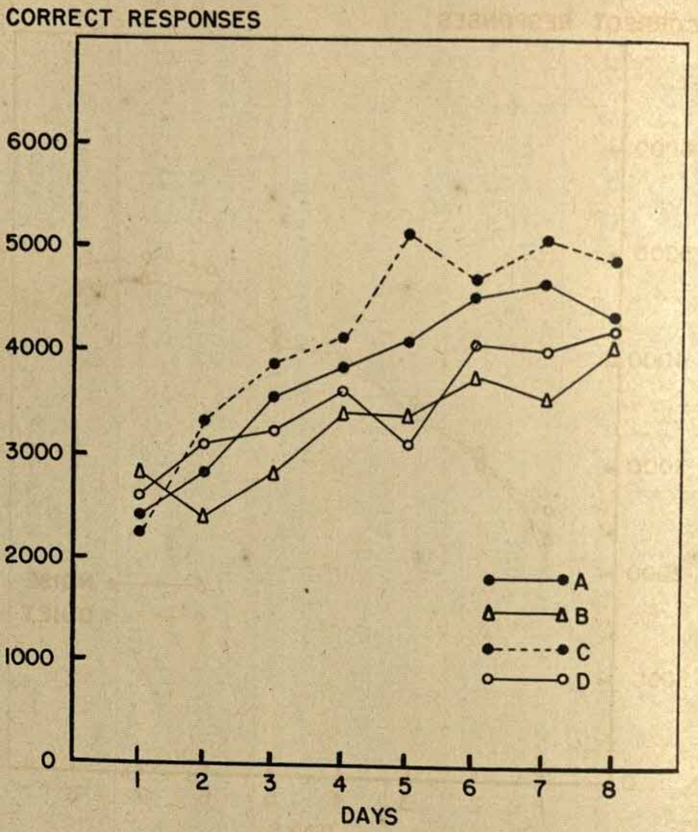


FIGURE 2
PERFORMANCE CURVES OF FOUR GROUPS OF *S*s UNDER QUIET FOR AN EIGHT-DAY PERIOD

spect to performance under noise and quiet conditions. Figures 3, 4, 5, and 6 present the performance curves of the four respective groups, i.e., the same *S*s working under both conditions. The curves for each group are based upon the data found in Table 1. A cross-over analysis of variance was computed for the four groups. Cochran and Cox (3) point out that in biological assay a design has been used which closely resembles the Latin square, but has some advantages when the number of treatments is small. In the simplest case, which we have here, there are two conditions, noise and quiet which have been alternated in *A B B A* sequence. The design and analysis are presented in Tables 4, 5, 6, and 7. The performance scores were converted to mean scores in the analyses of variance.

It should be observed that in Figure 6 there is a cross-over in performance

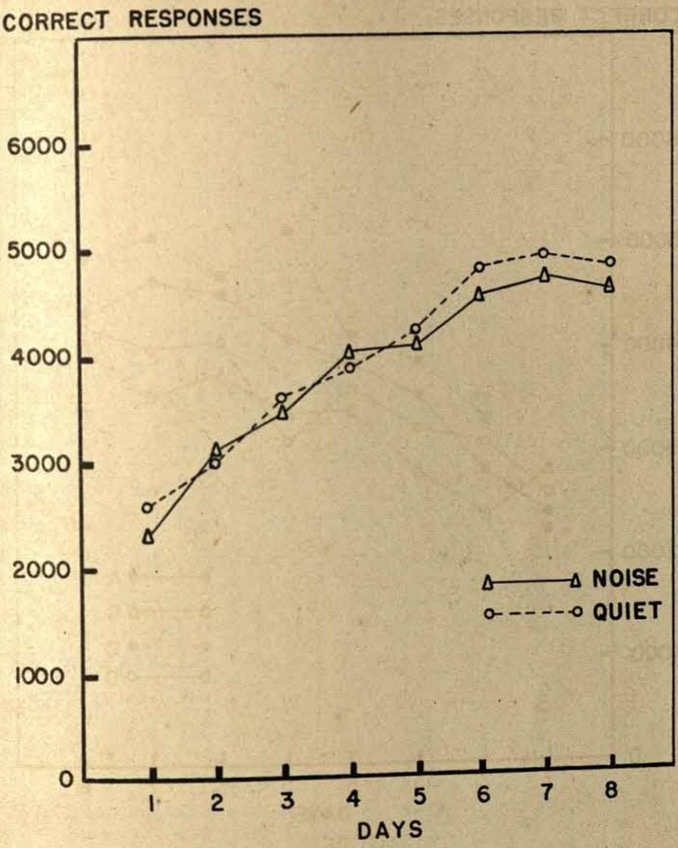


FIGURE 3
PERFORMANCE CURVES OF GROUP *A*: SAME *S*s WORKING UNDER NOISE AND QUIET EACH DAY

between Days 4 and 5, with Group *D* *S*s gradually increasing their response rate under noise conditions, when compared with their responses under the quiet condition. In theory, this cross-over was expected, and from the curves, Group *D* followed the predicted trend. Since the analysis of variance is a one-sided test, differences tend to disappear on both sides. It is clear from Figure 6 that differences on each side of the cross-over would balance out. To handle this problem it was decided to use a simple arithmetical artifact. The operations consisted of obtaining the differences between performance under noise from that under quiet for Days 5, 6, 7, and 8, and subtracting these differences for each of the cited days from the performance means under quiet. It should be noted that the empirical data

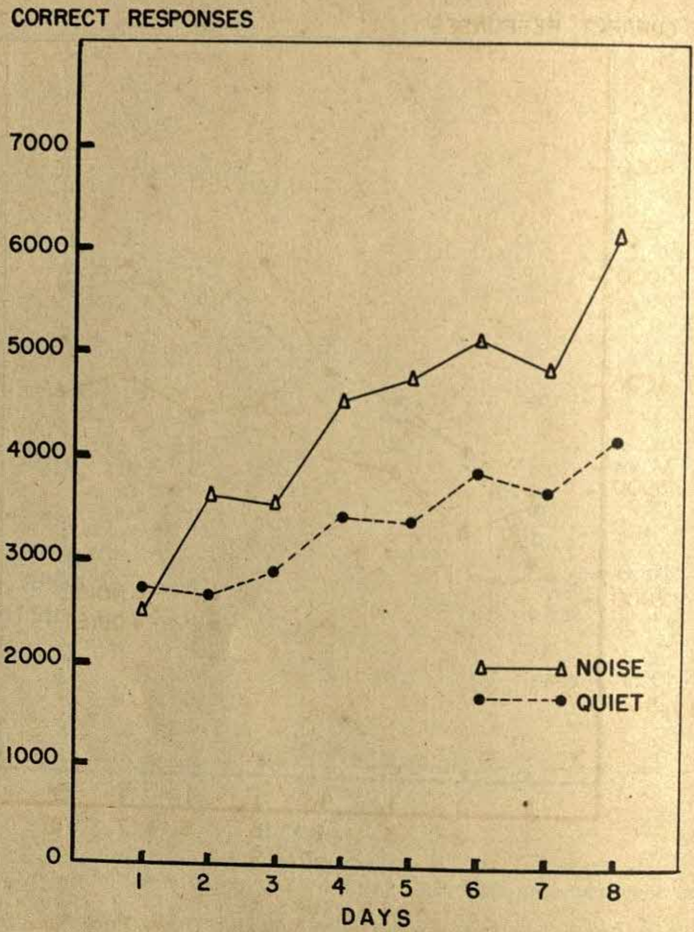


FIGURE 4
PERFORMANCE CURVES OF GROUP *B*: SAME *S*s WORKING UNDER BOTH NOISE AND QUIET EACH DAY

for Days 5, 6, 7, and 8 show Group *D* performing better under noise than under quiet conditions. The result of the transformation was to eliminate the cross-over and place the performance curve for noise the same distance between that of the quiet curve, as it originally was above it.

To summarize the results: (*a*) Group *B* subjects produced a significantly higher response rate under Noise conditions when compared with the three other groups; (*b*) Group *C* maintained a higher response rate under Quiet, but the difference was statistically significant at the .05 level on only two

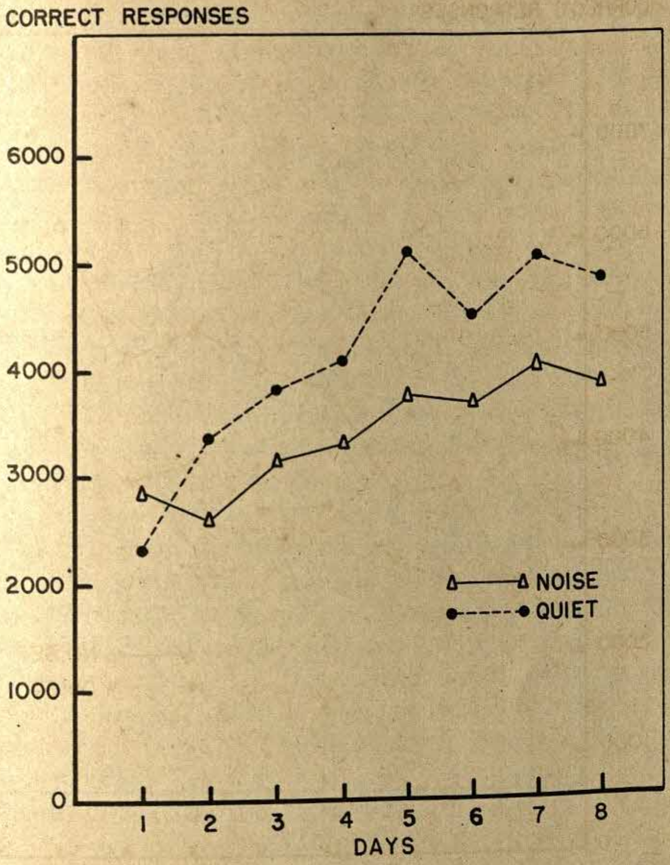


FIGURE 5
PERFORMANCE CURVES FOR GROUP *C*: SAME *S*s WORKING UNDER BOTH NOISE AND QUIET EACH SESSION

of the four covariance analyses; (*c*) Groups *B* and *C* performance deteriorated somewhat during the 72-hour interval between Days 5 and 6; (*d*) in comparing the intra-group trends Group *A* produced no significant differential response rate in either direction. Groups *B* and *C* produced statistically significant differential over-all response rates. Group *D* also produced a statistically significant differential response rate over the eight sessions. This difference was indicated after the arithmetical transformation had been completed in order to "eliminate" the cross-over.

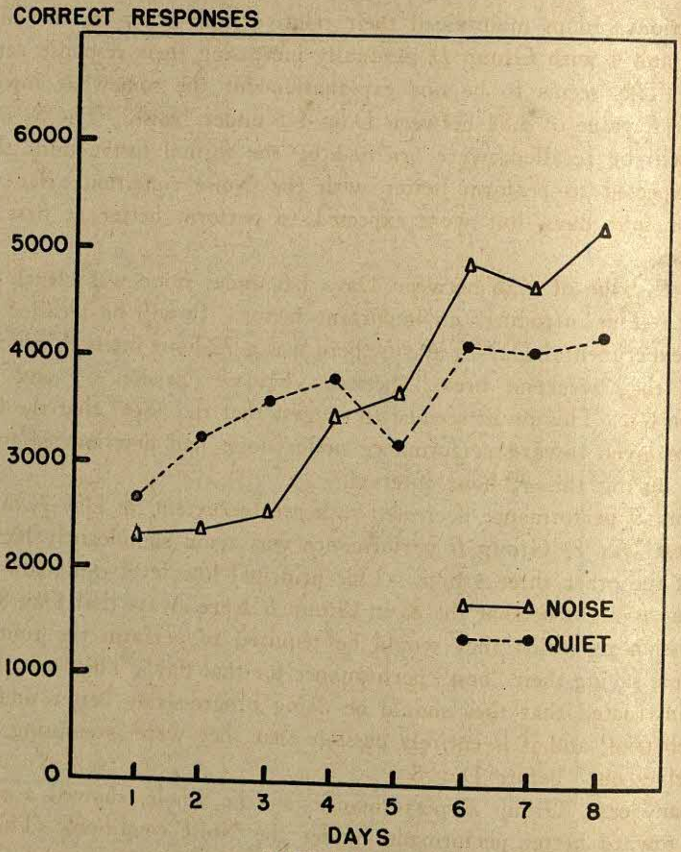


FIGURE 6
PERFORMANCE CURVES OF GROUP *D*: SAME *S*s WORKING UNDER BOTH NOISE AND QUIET EACH SESSION

D. DISCUSSION

Unfortunately, there is a lack of systematic data with which to compare the cited results. The reference here is to overt performance changes, and not to physiological measures such as *GSR*'s or action potential measures.

The intergroup analysis will be considered first. The instructions seemed to have an immediate effect upon the performance of the subjects. The various inversions of the groups from Day 1-2 under Noise was significant at the .05 level with an over-all *F* ratio of 7.87. Group *B* was clearly trying harder under the noise condition, while Group *C* appeared to be depressing their performance when the noise condition was introduced.

The various groups maintained their relative rank order on the curves for Days 3 and 4 with Group *D* gradually increasing their response rate under Noise. This seems to be one explanation for the somewhat lowered covariance *F* value of 3.81 between Days 1-5 under Noise. The *S*s in Group *D,* it will be recalled, were oriented by the formal instructions that they were expected to perform better with the Noise condition after the first three or four days, but were expected to perform better at first without the noise.

The *F* value of 1.26 between Days 1-6 under noise was clearly not significant. This introduces an important factor. It will be recalled that between experimental Days 5 and 6 there was a 72-hour interval which can be termed the "weekend break" between Friday (Session 5) and Monday (Session 6). This result seemed to suggest that the "Set" that the Group *B* *S*s were given toward performance under noise had deteriorated to a large degree during this 72-hour interval.

Group *B* performance decreased to a greater extent on Day 7, but on the final test Day 8, Group *B* performance was again significantly higher than that of the other three groups. One principal low level explanation of this occurrence might be that the *S*s in Group *B* were aware that Day 8 was the last session in which they would be required to perform the routine task, and were saving their "best" performance for that day. The *S*s in this group were instructed that they should be doing progressively better under Noise on each trial, and it is entirely possible that they were attempting to "burn themselves out" before Day 8.

In any case, Group *B* performance, on the whole, showed a significant trend toward better performance under the Noise condition. This is particularly evident when the work curves under both Noise and Quiet for Group *B* are compared. The analysis of variance in the cross-over design shows an *F* value for the "treatments" of 21.55 with a probability of less than .01. Figure 4 shows Group *B* as having performed slightly better under quiet conditions than under noise on Day 1 (initial measure). This difference, however, was not statistically significant. On Day 2 there was an inversion and Group *B* did consistently better with the noise. It is not untenable to suppose that these *S*s depressed their response rate with the record off. One female subject in this group reported that the reason for her better performance under noise was that she was "compelled" to add faster in order to keep from hearing the record.

The evaluation of such judgments is clearly not within the scope of this investigation. It is recognized that there is probably a multitudinous number of reasons that particular *S*s could give for their respective performances. However, in general, these could not be verified. Some typical reasons that were given for what some *S*s thought was a poor showing were: "I'm tired from studying"; "I have a test tomorrow, and can't think of anything else"; "I'm a pledge and my fraternity or sorority made me do such and such a thing today."

It is clear that the subjects used in the investigation were not in the same physiological condition from day to day. On one day a subject would appear cheerful and enthusiastic about the task; on another day the same subject would come into the experimental room, and give the impression that this was just another routine duty to perform. Clearly, these judgments are made from the experimenter's standpoint, and should not be interpreted as absolute in any sense. It is entirely possible that the investigator himself was not in the same physiological state on succeeding days.

Group *C* subjects produced a higher daily response rate under quiet conditions, but differences were statistically significant in only two of the four covariance analyses under the quiet conditions. As was pointed out with Group *B* under Noise, the Group *C* performance under Quiet appeared to deteriorate in the 72-hour interval between Days 5 and 6. This occurrence suggests that if instructions are to be an important variable in this particular type of work task, that frequent reinforcements with the particular set of instructions is desirable. If this condition is not fulfilled it would appear that extinction of the desired effect would soon take place.

An important phase of the analysis was embodied in the comparisons that were carried out between the four groups, separately, to determine whether each group showed any significant trend to perform differentially better under one condition than under the other. Group *A* were told only that they were serving in an experiment on the effects of noise. The analysis shows no significant trend in either direction in favor of noise or quiet. This group, as with the others, probably brought some bias into the experimental room toward one condition or the other. There was a slightly higher response rate for Group *A* under the Quiet conditions, but the difference was not significant. The implication here is that these subjects were not differentially disturbed by the noise condition.

Group *B* subjects, having been oriented toward a higher response rate under noise, produced a significant trend in that direction. Noise, here, certainly had no inhibiting effects. Group *C* produced a differentially significant trend toward better performance under the quiet condition; while Group *D*, which received the "complex" instructions, produced a significant trend toward better performance under quiet at first, with a gradual increase in response rate under noise toward the later sessions.

E. Conclusions

1. Formal instructions, or giving the subject an orientation as to what direction to work in, appears to be a significant variable in relation to routine production tasks. In order to produce differentially significant results, and eliminate vacillation of performance, the major property of "Set" seems to be that they be clear cut, and orient the subject in a precise direction.
2. Once the desired direction is produced, the evidence seems to indicate that frequent "Set" reinforcements are necessary to maintain superior performance in that direction, at least in relation to the routine verbal addition task employed.
3. Verbal Noise, *per se,* of a given intensity, does not appear to have any necessary effect upon the execution of routine work tasks. This finding seems compatible with some physiological results—that subjects gradually become adapted to the noise.
4. There is not consistent difference in production when no effort is made to give subjects formal instructions with respect to the condition under which they are expected to perform better.

References

1. Baker, K. H. Pre-experimental set in distraction experiments. *J. Gen. Psychol.,* 1937, **16**, 471-488.
2. Berrien, F. K. Effects of noise. *Psych. Bull.,* 1946, **43**, 141-146.
3. Cochran, W. G., & Cox, G. M. Experimental Designs. New York: Wiley, 1950. Pp. ix+454.
4. Davis, R. C. Electrical skin resistance before, during, and after a period of noise stimulation. *J. Exper. Psychol.,* 1932, **15**, 108-117.
5. ———. The muscular tension reflex and two of its modifying conditions. *Ind. Univ. Pub.,* Science Series, 1935, No. 3.
6. Edwards, A. L. Experimental Design in Psychological Research. New York: Rinehart, 1950. Pp. xiv+446.
7. Ford, A. Attention automatization: An investigation of the transitional nature of mind. *Am. J. Psychol.,* 1929, **41**, 1-32.
8. Freeman, G. L. Changes in tension pattern and total energy expenditure during adaptation to distracting stimuli. *Am. J. Psychol.,* 1939, **52**, 354-360.
9. Harmon, F. L. The effects of noise upon certain psychological and physiological processes. *Arch. of Psychol.,* 1933, No. 147.
10. Morgan, J. J. B. The overcoming of distraction and other resistances. *Arch. of Psychol.,* 1916, No. 35.
11. Ryan, T. A. Work and Effort: The Psychology of Production. New York: Ronald Press, 1947. Pp. vii+232.

12. STARCH, D. Experiments in Educational Psychology. New York: Macmillan, 1911.

13. VERNON, H. M., & WARNER, G. G. Objective and subjective tests for noise. *Personnel J.,* 1932-33, **11**, 141-149.

Institute of Educational Research
Indiana University
Bloomington, Indiana

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 299-306.

CERTAIN FACTORS IN PRODUCING COMPLEXITY OF RESPONSE TO A SINGLE PULSE OF LIGHT*

Department of Psychology, Michigan State College

S. HOWARD BARTLEY AND FRANK RAY WILKINSON

A. THE PROBLEM

It has long been known that a single brief pulse of light elicits more than a simple short-lived sensory response (4, 6, 8). The complexity has been divided into phases by several writers (3, 6). All but the first phase has been considered some sort of a series of after-affects, commonly called after-images. The implication has long been that these are of different origin than the first phase, called the primary sensation or "primary image." The findings of all workers have varied considerably depending upon general conditions. Early workers were not able to control these conditions precisely and this has precluded exact repetition of their work. Two general methods of experimentation were used: (*a*) a moving slit behind which was the source of brief illumination, and (*b*) a stationary brief target. These two methods produced somewhat different results. It is our purpose here to refer only to the findings obtained by the second method, for it would seem that the first method involves unnecessarily complicated stimulus conditions.

Dittler and Eisenmeier (3) enumerated eight phases in the complete response to brief stimulation. These are pictured in Figure 1. Judd (6) corroborated these stages in his study of Phase 5, the Purkinje after-image. For extensive listing of the men who worked in the problem of after-image production by brief stimulation, see Judd's article.

In 1942, Bartley (1) reported that a single pulse[1] of light, under some conditions, gives rise to two instead of one flash, without any very prominent accompanying phenomena. He also reported a large secondary wave appearing in the optic nerve discharge under similar stimulus conditions in a series of neurophysiological experiments. This wave was taken to be the neural basis for the second flash. In the original investigation, Bartley found that a very weak pulse produced a single flash and a certain range of more in-

---

*Received in the Editorial Office on November 22, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The term *pulse* is substituted for *flash* so as to avoid using the same word for both stimulus and response.

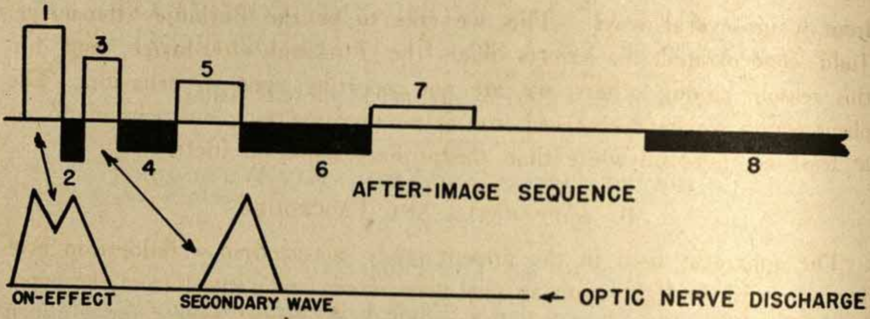


FIGURE 1
Upper, the response to a short pulse of light consisting at most in about eight phases: (1) The primary image; (2) a short negative phase; (3) the first positive phase or Hering after-image; (4) a second negative after-image; (5) the second positive or Purkinje after-image; (6) a third negative after-image; (7) a third positive after-image; and (8) a fourth negative after-image. Each phase is probably a little longer than its predecessor of the same sign. Lower, schema showing on-effect and secondary wave of optic nerve discharge.

tense or longer pulses produced the double flash experience. Still higher intensities, or still longer pulses again produced a single flash. Thus two transitions were involved; i.e., (*a*) going from the conditions producing one flash to those producing two flashes, and (*b*) going from conditions producing two flashes to those producing one as the variable in question was increased. The analysis of the stimulus conditions was not extended beyond a preliminary manipulation of intensity and duration of pulse.

The present article reports the attempt to study the two-flash phenomenon further by the manipulation of pulse intensity and duration and by varying the area of the image on the retina.

Examination of Figure 1 will show that the first two positive phases following the primary sensation are separated by a short negative phase. Judd found that, under some conditions, these two positive phases were observed as only a single phase. Either the two literally became one or else the negative phase was so trivial as to be overlooked. This finding explained some of the discrepancies between the reports of some of the workers. At no time was the first positive phase (Phase 3, or Hering's after-image) found to vary in color quality, or even to be anything but achromatic. The second flash we are reporting is the Hering after-image, or the combination of the Hering after-image and the second positive phase (Phase 5, the Purkinje after-image) if the two perhaps coalesced under appropriate conditions. Some of our incidental observations showed that under certain conditions a colored after-image followed the second flash and was very distinguishable from it in several ways. This we take to be the Purkinje after-image. Judd concentrated his efforts upon the Purkinje after-image, and for this reason, among others, we are not reporting upon its behavior. The phenomenon we have observed was at no time anything but achromatic, or at least no more off-white than the primary sensation itself.

B. Apparatus and Procedure

The apparatus used in the present study consisted in a Balopticon projecting a disk of light upon an opal-glass screen and viewed from the opposite side. The screen was covered by black cardboard, except for a 6-inch disk-shaped opening as the stimulus object. The duration of the pulses was varied by an Abney episcotister whose open sector could be enlarged or reduced during disk rotation.

The part of the investigation to be reported upon here consisted in setting pulse duration at each of 10 different values and varying the intensity upward so that the second flash just became apparent, and further increasing intensity so that the second flash just disappeared.

Six different stimulus areas were used. These, however, were obtained, not by varying target size but by varying distance from the target. Figure 2 indicates these distances as 1.5, 2, 3, 4, 6, and 15 feet. This figure also gives the populations of the rods and cones in relation to the target.

These distances, in turn, provided for retinal stimulus areas subtending approximately 19, 14.4, 9.5, 7.2, 5.7 and slightly less than 2 degrees overall. The rate of presentation was one pulse every .86 second. The durations of the pulses ranged from about 3 msec. to 142 msec.

Three observers in addition to the authors were used in the first part of the investigation. For the data to be reported, only the authors were involved. The material in Figure 3 is from one of these. The data from the other were substantially the same in kind.

C. Results

Figure 3 shows our experimental results. The ordinate represents pulse intensity in the logarithm of candles per square foot, and the abscissa, pulse duration in milliseconds. There are two curves for each viewing distance used. It will be noted that no curves are presented for the 15-foot distance. No second flash observations were obtainable at this distance. The goodness (distinctness and general impressiveness) of the second flash diminished as the viewing distance was increased and the area covered by the retinal image was thus reduced.

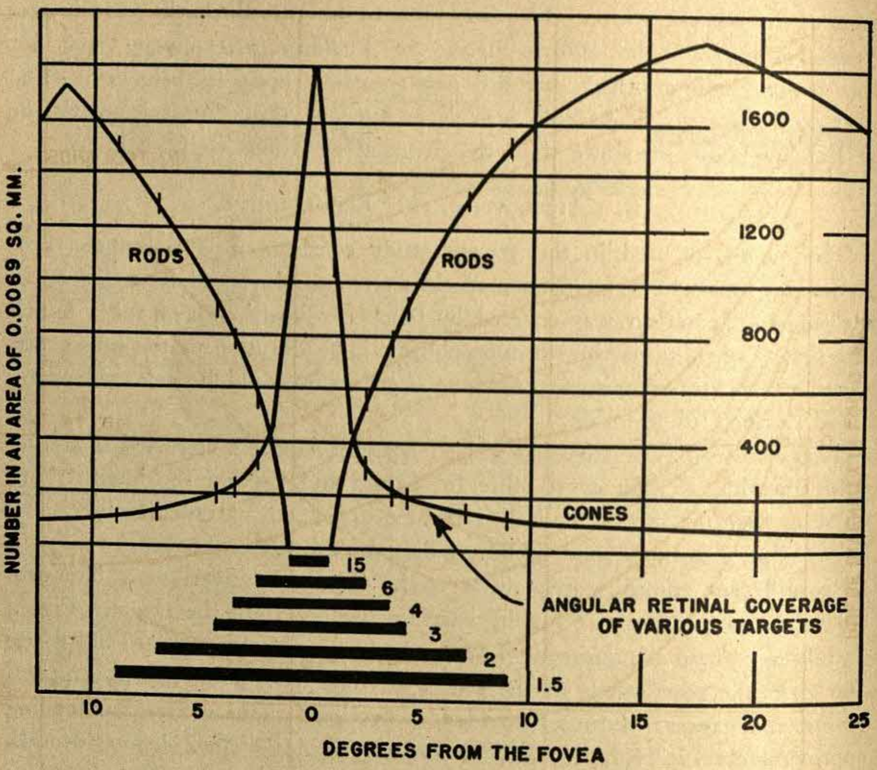


FIGURE 2
SHOWING DISTRIBUTION OF RODS AND CONES ON HUMAN RETINA AFTER OSTERBERG, AND THE RETINAL COVERAGE OF THE SEVERAL VISUAL TARGETS USED

It will be seen that the pair of curves for the 1.5-foot viewing distance (largest retinal area) lies lowest in the intensity range, and that an orderly sequence is followed in the shift upward in the intensity range as viewing distance is increased. This holds for pulse durations up to about 70 msec. where the one-flash to two-flash transition curves for the 4-foot and 6-foot viewing distances reverse, and rise as pulse duration is further increased. A beginning rise is suggested in the bottom curves for the 2-foot and 3-foot viewing distances. Some rise would be expected for the 1.5-foot curve, but it would not occur except at pulse durations, too long to be shown in the graph. Ideally, the bottom curve and top curve for each viewing distance should come together both at very long and very short pulse durations, so as to form a closed loop. Inside the loop is the family of intensity-dura-

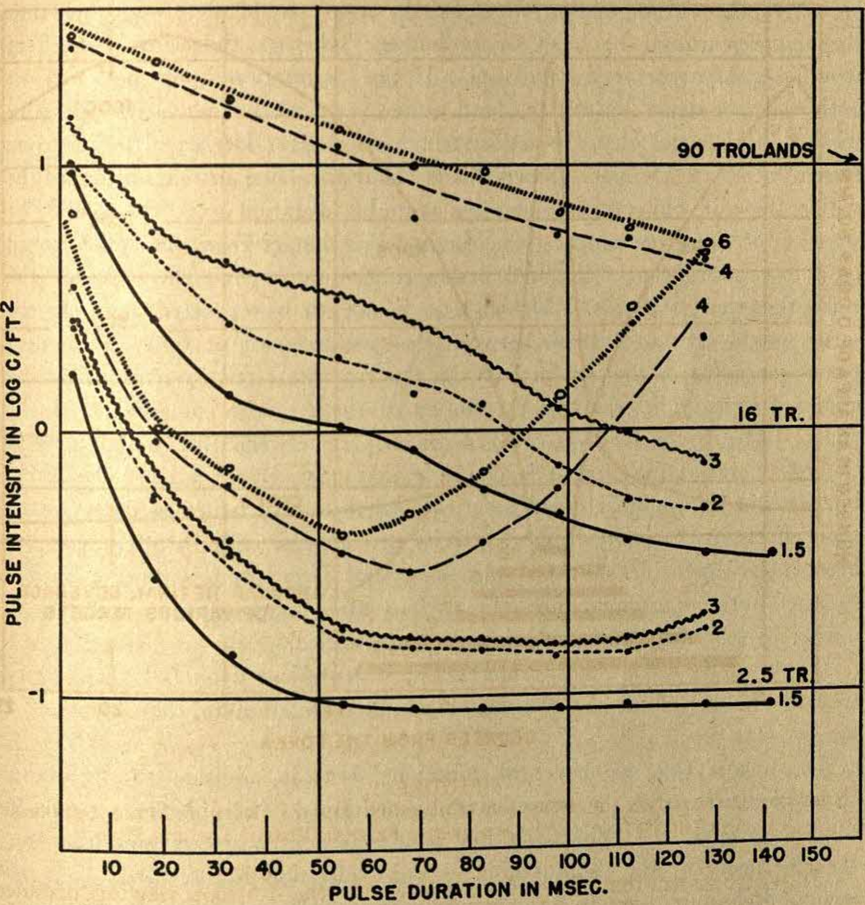


FIGURE 3
SHOWING EXPERIMENTAL RESULTS (SEE TEXT FOR EXPLANATION)

tion combinations of pulse eliciting the two-flash phenomenon and outside the combinations eliciting only the "primary" sensation.

It may be significant that the major slope of the bottom curve in each pair is concluded by the time the pulse duration reaches about 60 msec. It is in this region that the second flash is most prominent. There is no doubt but that the second flash is dependent upon the use of a limited time and intensity range in stimulation. Manipulation of these two factors would be expected to involve the following consequences. Manipulation of intensity would determine whether only rods, both rods and cones, or only cones would be activated. Manipulation of pulse duration would determine whether the response would be brief or prolonged; whether the neural off-effect would be absent or well developed. If the stimulation is in the range of both rod and cone sensitivity, both systems would be activated. Earlier work (7) indicates that the rod system has a longer latency than the cone system by about 55 msec. Since this is about the same in magnitude as the major slope in our curves along the stimulus duration axis, there may be some connection between the two. Light pulses longer than this value would tend to produce cone originated flashes continuing beyond the onset of the rod originated responses. Although it is not to be expected that an observer could ever distinguish between the two in point of time, the neural consequences of setting up activity in the two sense-cell systems might be expected to vary in terms of the timing of the response components of the two. There is considerable convergence in and between the two systems in the peripheral retina, and it is plausible that part, if not all, of the after-image sequence might be based upon the elaboration of neural activity rather than effects in the photochemical activity of the sense cells themselves as has been classically held. Such an idea as neural interaction is more plausible in attempting to account for why various after-images appear and fail to appear as stimulus timing and intensity are varied.

Whatever the second flash represents, it is dependent upon (*a*) the activity of both systems, and (*b*) a peculiar temporal relation in the behavior of the two systems.

To support this, we have the following facts in addition to the major slope terminations of the curves already mentioned: (*a*) The curves finally tend to reverse, thus signifying the requirement at a greater intensity for the emergence of the second flash, as pulse duration is increased; (*b*) the second flash diminishes in prominence as intensities are raised much above the level of their first emergence; (*c*) no second flash is seen when the image of the pulse falls entirely within the fovea; (*d*) the prominence of the second flash is increased as the area of the retina covered by the stimulus is increased. This increased prominence cannot be brought about by increasing intensity beyond a small amount. Further definitive features of the outcome are (*e*) that the second flash is always the same hue as the first, at least when the stimulus is white or nearly so, and (*f*) that the second flash cannot be elicited at scotopic intensity levels. This is in contrast to the behavior of the Purkinje after-image or phase 5 in the series described in Figure 1, which varies in hue as a function of stimulus intensity.

To continue the consideration of the possible origin of the second flash, the findings (1, 2) on the optic nerve discharge must be re-examined. In the optic nerve discharge, Bartley (1) found that an extra wave appeared when the stimulus pulse was in the mid-range of intensity or duration. He took this to be the retinal origin of the second flash. While we can hold to that view, this does not in itself tell whether the second flash is a part of the sequence of events in the rod pathway, or in the cone pathway. It is difficult to find a satisfactory statement regarding the histology of the rabbit eye. Hence the kind of helpful information that would come from histology will have to await a specific study motivated by such purposes as ours. The record of the optic nerve discharge does contain evidence that the on-effect is duplex. With weak stimulation, it is but a single wave. As intensity is increased a second wave becomes part of the on-effect, and this wave appears to emerge ahead of the original wave. This would be in line with expectations were the rabbit to have two diverse sense-cell populations. The initial wave would seem to represent the effects of sense cells of high sensitivity, but longer latency, and the second wave emerging in front to represent sense cells of lower sensitivity coming into play only as intensity is raised. The finding would fit the older deductions that the rods have a greater latency than the cones. In the records, the delay between the first and second components of the on-effect is of the same order as that deduced by McDougall (7) for the human.

The secondary wave in the record of the optic nerve discharge arises only after 150 or more msec. following the onset of the first component in the on-effect. Under some conditions, the secondary wave may arise much later than this. When the stimulus pulses are long (let's say 200 msec.) and a definite off-effect is recorded, a vestige of the secondary wave follows the off-effect, or at about 250 msec. after the beginning of the on-effect. Thus it is that we cannot say that the secondary wave represents an initial response of either the cones or rods. It is not the origin of a "primary sensation" but of truly an "after-image," as the older men called all of the effects following the first flash. Since the second flash arises only at photopic levels and only while rods are yet in action, it must arise as an integrated effect from both pathways. This is not strange since in the human eye certain of the rods and cones are known to converge on the same bipolar cells. If then, the secondary wave in the optic nerve discharge is the retinal origin of the second flash, we can believe that the secondary wave is a result of converged activity of two kinds of receptor populations. This can be held in spite of the probability that the rabbit, as an animal, does not respond overtly to spectral differences. There is a decided distinction between "color vision" and the presence or absence of an adequate receptor mechanism for overt differential behavior to spectral stimulation, or to two ranges of the intensity continuum.

A tentative suggestion can be made regarding the after-image series pictured in Figure 1. Since there was only one large wave following the on-effect, it is plausible that only part of the after-image series originates in the retina. Since we have pretty surely identified this with the second of our two flashes and Phase 3, or the Hering after-image, it may be that the phases following this are centrally initiated. It must be kept in mind that such deductions are not based upon physiological findings from the human subject. They would hold only if the human and rabbit eyes are sufficiently similar. Such a comparison would not be based upon the assumption that rabbits see any after-images at all.

REFERENCES

1. BARTLEY, S. H. The features of the optic nerve discharge underlying recurrent vision. *J. Exper. Psychol.*, 1942, **30**, 125-135.
2. BARTLEY, S. H., & BISHOP, G. H. Some features of the optic nerve discharge in the rabbit and cat. *J. Cell. & Comp. Physiol.*, 1942, **19**, 79-93.
3. DITTLER, R., & EISENMEIER, J. Ueber des erste positive Nachbild nach kurzdauerender Reizung des Sehorgans Mittelst bewegter Lichtquelle. *Arch. f. d. ges. Physiol.*, 1909, **126**, 610-647.
4. FRÖHLICH, F. W. Untersuchungen über periodische Nachbilder. *Zsch. f. Sinnesphysiol.*, 1921, **52**, 60-88.

4*a*. ———. Über den Einfluss der Hell und Dunkeladaptation auf den verlauf der periodischen Nachbilder. *Zsch. f. Sinnesphysiol.*, 1922, **53**, 79-107.

4*b*. ———. Über die Abhängigheit der periodischen Nachbilder von der Dauer der Belichtung. *Zsch. f. Sinnesphysiol.*, 1922, **53**, 108-121.

5. HERING, E. Eine Methode zur Beobachtung und Zeitbestimmung der erstes positiven Nachbildes kleiner bewegte objecte. *Arch. f. d. ges. Physiol.*, 1909, **126**, 604.
6. JUDD, D. B. A quantitative investigation of the Purkinje after-image. *Am. J. Psychol.*, 1927, **38**, 507-533.
7. MCDOUGALL, W. The sensations excited by a single momentary stimulation of the eye. *Brit. J. Psychol.*, 1904, **1**, 78-113.
8. PURKINJE, J. Beobachtung u. Versuche zur Physiol d. Sinne. Beiträge zur Kenntnis des Sehens in Subjective. *Hinsicht*, 1823, **1**, 97.

Department of Psychology
Michigan State College
East Lansing, Michigan

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 307-312.

QUANTITATIVE STUDIES OF VERBAL REINFORCEMENT IN CLASSROOM SITUATIONS: I. DIFFERENTIAL REINFORCEMENT RELATED TO THE FREQUENCY OF ERROR AND CORRECT RESPONSES*

Institute of Educational Research, Indiana University

DONAVON AUBLE AND E. VICTOR MECH[1]

A. INTRODUCTION

During the last three decades, several studies have been reported dealing with the rôle of such motivational techniques as praise and reproof in classroom situations. Data reported by Schmidt (4) and Hurlock (2) indicate that praise is superior to reproof in producing positive performance changes. As a consequence, many educational psychologists have tended to identify the "law of effect" with improvement in performance and have neglected to focus sufficient attention upon performance that reflects a high frequency of error.

With the meager empirical data available regarding classroom experimentation, educational psychologists are not justified at present in "applying" any theoretical learning position to education nor in comparing their results with any particular learning positions such as those of Hull, Guthrie, Tolman, or Skinner, although Stroud (5) has indicated the need for such an approach.

The experiment being reported closely parallels the procedure described by Hurlock and, except for a few modifications, a general attempt at replication has been made. It appeared to the authors that in attempting to assess the relative efficacy of various incentives, simultaneously, Hurlock was increasing the complexity of an already complex task.

The procedure Hurlock used was to administer three treatments, praise, ignore, and reprove. The "control" group was part of the same class, but on the testing sessions left the other three groups and went to another room.

---

*Received in the Editorial Office on November 23, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The authors are grateful for the wholehearted coöperation and interest of Mr. Homer Terrell, principal of Central Elementary School, Martinsville, Indiana, in granting permission for a series of classroom experiments. The writers are indebted, also, to the teachers, Miss Marlene Sullivan, who handled the differential reinforcements received by Groups I and II, and Mrs. Jeanne Masters who administered the task to the control group.

Such a complex study increases the problem of objective interpretation. The "ignored" group, indeed, was not ignored—but was in the presence of both reinforcing and negative stimuli. It is also difficult to determine what the effects were on the control group belonging to the same class and then being placed in another room each day to be tested. The present experiment did not deal with "reproof," and attempted to reduce the complex situation described by Hurlock to a level that does not defy interpretation.

B. Design and Procedure

The 51 *S*s, third-grade pupils at Central Elementary School in Martinsville, Indiana, were divided as indicated in Table 1 into three subgroups.

TABLE 1
SUMMARY OF EXPERIMENTAL PROCEDURES

Group	*N*	1	Sessions 2-5
I Reinforcement	17	Pretest measure for the three groups	Received verbal reinforcement each session.
II Omit reinforcement	17		Verbal reinforcement omitted but were in presence of reinforcing stimuli received by Group I.
III Control	17		These *S*s were told nothing each day except that they were to work problems.

Group I, which was designated as the verbal reinforcement group, consisted of one-half of a third-grade class. All *S*s belonging to this group were randomly assigned. The remaining one-half of the same class were automatically placed in Group II, in which verbal reinforcement was omitted. Group III designated as the control, was composed of a separate third-grade class located in another part of the building, which simply did the required task each session without any explanation from their teacher. The initial test conditions were the same for all three groups. On Sessions 2-5 before the test papers were distributed, the names of the *S*s in the reinforcement group were called and the *S*s were asked to come to the front of the room and face the rest of the class. The teacher verbally reinforced them for the excellency of their work of the preceding day, as shown both in improvement made and general superiority over the other members of the class. They were encouraged to do better, to try to avoid any careless mistakes, and to add as many problems as the time permitted. Group III, a separate third-grade class in another section of the building, was given the same task for five sessions by their regular teacher and simply were asked to do the problems. The task was a set of addition and subtraction problems consisting mainly of one-digit number combinations. All *S*s were allowed six minutes a session to work on the problems. A sufficient number of problems were given each day so that the most rapid *S*s were not able to finish in the allotted time.

C. RESULTS

The transformed mean correct responses for the three groups appear in Figure 1. Group III (Control) gave a higher number of mean correct responses each day than did Group I (Reinforcement) or Group II (Omit reinforcement). An over-all analysis of variance described by Kogan (3) was computed to decide whether to reject the null hypothesis that the groups exhibit no difference in general trend. Due to the successive variances tending to be proportional to the successive means a square root transformation ($\sqrt{X}$) suggested by Bartlett (1), was carried out prior to making the analysis for trend. Table 2 presents a condensation of the analysis. The *F* of 2.35, significant at the .05 level, indicates that the Control group performed differentially better (as indicated by number of correct responses) than either reinforcement or omit-reinforcement groups.

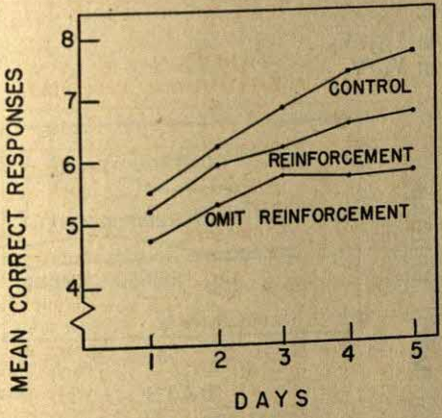


FIGURE 1

MEAN TRANSFORMED ($\sqrt{X}$) SCORES OF CORRECT RESPONSES MADE BY THREE GROUPS OF THIRD-GRADE *S*s IN A ROUTINE ARITHMETICAL TASK FOR FIVE DAYS

A similar analysis was carried out for the error responses shown in Figure 2. Prior to the application of the analysis of variance, a transformation of the data ($\sqrt{X + .5}$) was used to reduce heterogeneity of variance; after this transformation, the means and variances were no longer proportional and the heterogeneity of variance was markedly reduced.

TABLE 2
ANALYSIS OF VARIANCE OF THE $\sqrt{X}$ TRANSFORMED CORRECT RESPONSES

Source of variation	Sum of squares	*df*	Variance estimate	*F*	*P*
Methods	69.22	2	V_M	2.98^a	$> .05$
Trials	23.76	3	V_T		
MXT Interactions	5.80	6	V_{MT}	2.35^b	$< .05$
Component error	616.59	192	V_C		
Inter subject	557.34	48	V_S		
Intra subject	59.25	144	V_I	84.12^c	$< .001$
Total	715.37	203			

[a] Obtained by testing V_M/V_S.
[b] Obtained by testing V_{MT}/V_I.
[c] Obtained by testing V_M/V_I.

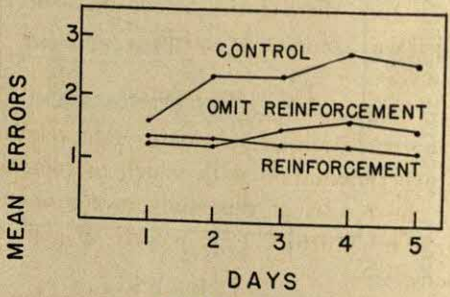


FIGURE 2
MEAN TRANSFORMED ($\sqrt{X + .5}$) SCORES OF ERROR RESPONSES MADE BY THREE GROUPS OF THIRD-GRADE *S*s IN A ROUTINE TASK FOR FIVE DAYS

The complete analysis of variance on the transformed scores for the three groups is shown in Table 3. The *F* of 5.91 is significant at the .01 level, and indicates that Groups I and II produced fewer mean errors than Group III.

A note of caution should be injected at this point with respect to the homogeneity of groups. An analysis of variance was computed for the initial measures on both error and correct responses, the *F*-ratios indicating that at the outset, the groups were drawn from the same population. However, the condition of homogeneity of variance was not satisfied for the group curves after the treatments were administered. Although the square root transformation markedly reduced this heterogeneity of variance it did not reduce it sufficiently to permit full confidence in the interpretation. This was especially evident in the error curves, and the obtained *F*-ratios are

TABLE 3

ANALYSIS OF VARIANCE OF THE $\sqrt{X + .5}$ TRANSFORMED ERROR RESPONSE

Source of variation	Sum of squares	*df*	Variance estimate	*F*	*P*
Methods	63.03	2	V_M	5.91[a]	$< .01$
Trials	1.65	3	V_T		
MXT Interactions	1.97	6	V_{MT}	1.06[b]	$> .05$
Component error	300.77	192	V_C		
Inter subject	255.98	98	V_S		
Intra subject	44.79	144	V_I	101.32[c]	$< .001$
Total	367.42	203			

[a] Obtained by testing V_M/V_S.

[b] Obtained by testing V_{MT}/V_I.

[c] Obtained by testing V_M/V_I.

probably somewhat higher than if the homogeneity of variance assumption underlying the model was completely satisfied.

D. DISCUSSION

In general, there is a dearth of systematic data relating verbal reinforcement to classroom experimentation with which to compare the cited results.

However, taking the results of this study in conjunction with those of a previous investigation by Hurlock (2), we are in a position to make some tentative comparisons.

Hurlock's results indicated that her "praised" group did the best work and the control group did the poorest. The data in the present experiment indicate that the control group showed a statistically significant trend for number of correct responses when compared with the differential reinforcement groups. Our results, then, are not compatible with those of Hurlock, and cast immediate doubt upon such vague generalizations as to "praise is better than to ignore." Actually, Hurlock reported combined data on *S*s in the fourth, fifth, and sixth grades. Such confounding makes a close analysis of her results extremely difficult.

If one were prone to terminate the analysis with the correct response curve, much interesting data would be obscured. In comparing the Figure 1 correct response curves with the Figure 2 error curves, it is evident that although the control group gave more correct responses, they also responded with more errors, and in turn, Group I (Reinforcement) had the fewest errors each day.

Analysis of the data appears to indicate:

1. Verbal reinforcement has the property of reducing variability of per-

formance in the routine task specified (although the group receiving reinforcement did not exceed the control group in correct responses, it did make significantly fewer errors).

2. There appears to exist some mediation of the reinforcing stimulus to the *S*s who did not directly receive it, but who were in the presence of the reinforcement. (There were no statistically significant differences between the differential reinforcement groups when compared as to frequency of error and correct responses, although the verbal reinforcement group responded with more correct and fewer error responses).

E. Summary

An experiment employing 51 third-grade *S*s was conducted in order to test the effects of differential verbal reinforcement upon routine performance in a classroom situation. The 17 *S*s in Group I were reinforced verbally for five successive days. Group II (17 *S*s) was in the presence of the reinforcing stimuli, but received no recognition from the teacher for their daily performance. Thus Groups I and II were members of the same class, while Group III (Control) was a separate third-grade class with their own teacher, and simply were asked to do the task daily.

The results showed: (*a*) Verbal reinforcement has the property of reducing the variability of performance in the routine task specified. (*b*) There appears to exist some positive mediation of the reinforcing verbal stimuli to *S*s who did not directly receive it, but who were in the presence of the reinforcement.

The results are contrasted with those of Hurlock, and are interpreted as indicating a general need for classroom experiments that are relatively simple in design but amenable to objective interpretation.

References

1. Bartlett, M. S. Square-root transformations in analysis of variance. *J. R. Statis. Soc. Suppl.*, 1936, **3**, 68-78.
2. Hurlock, E. B. An evaluation of certain incentives used in school work. *J. Educ. Psychol.*, 1925, **16**, 145-159.
3. Kogan, L. S. Analysis of variance—repeated measurements. *Psychol. Bull.*, 1948, **45**, 135-143.
4. Schmidt, H. O. The effects of praise and blame as incentives to learning. *Psychol. Monog.*, 1941, **53**, No. 240.
5. Stroud, J. B. Educational Psychology. *Annu. Rev. Psychol.*, 1951, Vol. II (C. P. Stone, *Ed.*).

Institute of Educational Research
Indiana University
Bloomington, Indiana

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 313-328.

HYPNOSIS AS EMOTIONAL RESPONSE: A THEORETICAL APPROACH*[1]

Department of Psychology, Long Island University

HENRY GUZE

Theories of hypnosis, historically, have accounted for the hypnotic state as a sleep state (48, 56), or as some similar passive state during which the brain is inhibited at the cortical level. Pavlov (48) places hypnotic response on a continuum with waking behavior and sleep, pointing out that the hypnotic "phases appear to us on the one hand, as different degrees of the extent of inhibition, i.e., a more or less spreading of inhibition in the areas of the hemispheres themselves, and also in different parts of the brain; and, on the other hand, as different degrees of intensity of inhibition in the form of varying depths of inhibition at one and the same place. "Naturally," says Pavlov, "with the greater complexity of the human brain the hypnotic phenomena are more varied in the human than in the animal." Pavlov continues to conclude that schizophrenic symptoms are the expression of a chronic hypnotic state. Recent suggestion has been made by Schilder and Kauders (56) that certain pathological conditions may be spontaneous manifestations of the hypnotic state. They assume that the sleep center in the third ventricle is involved in hypnotic activity and that such activity through the area of the third ventricle and the hypothalamic area acts to influence the vegetative functions of the organism. In this regard it is important to note that the hypothalamic area has been considered as an integrative center for the emotions (2). Furthermore, Bateman and Papez (3) point out that very severe destruction of hypothalamic cells was found in most of the brains of markedly deteriorated (psychotic) patients who had lived an emotionally non-reactive life in the hospital. What the foregoing data indicates, it is not easy to conclude. However, there seems to be some suggestion of a mechanism that may be similar for sleep, emotion, hypnosis, and psychosis. It has been postulated by Nauta (47) that there must be a center for sleep in the hypothalamic region, his statement being based upon animal research. At least it can be assumed that the hypothalamus is involved in the sleep-waking organization.

---

*Received in the Editorial Office on November 28, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Paper delivered at second annual meeting of the Society for Clinical and Experimental Hypnosis, New York Academy of Sciences, September 29, 1951.

Whether or not it is justified to regard hypnosis and sleep as closely related is considerably open to question. Such challenge, of course, would cast doubt on the provocative hypothesis outlined in the foregoing paragraph. From this standpoint, Gorton (22), in his excellent review, has pointed out that the evidence relating sleep to hypnosis is not convincing, and various studies seem to show that hypnosis and the waking state are quite similar in terms of electroencephalography, as demonstrated by Loomis, Harvey, and Hobart (42). Of course it must be remembered that Pavlov (47) placed hypnosis on the continuum between the waking state and complete sleep, thus the question may arise as to whether the lack of electroencephalographic changes typical of sleep entirely disproves any relationship between the two.

Dynamic or analytic theories of hypnosis have attempted to recognize some drive characteristic of the hypnotic state. Again this does not necessarily remove it from the sleep-waking continuum, since sleep itself has drive characteristics, and reduces the cortical control over the emotions, particularly when the subject is very sleepy but must keep awake. That a drive aspect is prominent in hypnotic behavior is dramatically manifest by the vast array of classic phenomena of the state. Schilder (55) sees in hypnosis, expressions of libidinal response, and is even aware of the rôle of the hypnotist from this standpoint. Says he, "We must be mindful of the fact that the sexual constitution of the individual finds expression in the nature and form of the hypnosis." Jones (35) too, sees hypnotic behavior as libidinal but stresses the aspect of narcissism. Freud (19) sees in hypnotic behavior the element of subjection. It is similar to being in love. There is the paralysis, also derived from someone with superior power and someone who is without power. He elaborates further in terms of group psychology and the primal horde. All of the foregoing theories outlined above seem to have a single characteristic in common, that is a general change in the emotional reactivity of the individual. Similarly the theories of MacDougall (43) who introduces an instinct of submission, and that of Rivers (51) who relates hypnotic and group phenomena to herd behavior, seem to imply that hypnotic phenomena are somehow related to various aspects of emotional response.

Kubie and Margolin (39) regard hypnotic induction as a partial sleep wherein there is blurring of the distinction of the ego and the outside world due to decrease in environmental sensory stimulation. They assume that finally there is an incorporation of the hypnotist into the subject. To them hypnosis is an extension of maximal attention in terms of a single field of sensation. This point of view is much like that of Freud (19), who, however emphasizes attention to the hypnotist involving withdrawal from the outside world. This, he believes, constitutes the similarity between sleep and hypnosis. It is to be noticed that while the various theoreticians describe the hypnotic phenomenon in its multiplicity of aspects, they are much like the six blind men of John Godfrey Saxe, each touching a different part of hypnotic behavior, but not arriving at the basic underlying characteristic. This is, of course, understandable in view of the great complexity of the problem, and the present writer can only hope, without certainty, that he has eluded this pitfall.

There are, of course, other approaches aside from those mentioned above, such as the views of Bernheim (4), Braid (6), Charcot (14), and others. Recent times have seen the theory of White (67) who has proposed that a functional decortication of a kind alters the consciousness of the subject and makes him more prone to hypnotic striving. This theory is a dynamic one wherein it is assumed that the subject strives to act like a person who is hypnotized (or what he conceives of as a hypnotized person). But White (67) fails to explain why such striving occurs, attributing it to aptitude. Thus differences in hypnotic susceptibility become differences in aptitude for hypnotic striving, a somewhat vague picture indeed. Why should one strive in this area at all? This view is not unlike the Sarbin (53) theory that hypnosis is a form of (or related to) dramatic aptitude. While a relationship is possible, hypnotic states are much less voluntary and thus the concept of aptitude is suspect. Sarbin himself is aware of the problem in this area, although it must not be assumed that lack of full voluntary action negates the possibility of aptitude. However, aptitude in this case is not a clear term. Is Sarbin referring to "rôle-taking" or rôle-taking at a lower (or lesser) level of consciousness? It may even be questioned as to whether hypnotizability can be regarded as a unit trait of behavior or a complex of trait function.

Sarbin says that the person in hypnosis strives to take the rôle of the hypnotized person, a very circular statement. The success of his striving is a function of a favorable motivation he would claim, also a favorable rôle perception, and rôle taking aptitude. A congruence between the subject's self concept and rôle of hypnotic subject is to be seen as favorable motivation. Rôle taking aptitude is, he claims, a function of the subject's ability to use "as-if" formulations.

In experimental and therapeutic sessions the pressures placed on a subject would seem to negate much of the emphasis which Sarbin places upon cooperative voluntary participation in a game-like affair. This writer and his co-worker, Kline (36), have observed subjects under stress situations in hypnosis which seemed to place severe tax upon them in behavior that could hardly be called voluntary. This writer (24, 25) believes that subjects (of a certain nature) can be forced into hypnotic states and forced to behave in a way which is completely inimical to their morals or attitudes. This, of course, depends on the personality of the subject in terms of impulse-handling in every day life, and will be discussed later as part of the current theoretical formulation.

Researches by Schneck (57), Kline (37), and Guze (24) have demonstrated an array of hypnotic responses which seem to have diagnostic significance. They are essentially ways in which a subject handles the hypnotic state and its concomitants.

It is the thesis of the present paper that hypnosis is a dynamic state rather than a passive one and that all hypnotic responses are significant as indicators of the particular personality organization of the subject. The hypnotic state is in itself an altered functional state of an undefined emotional nature, being marked by fear, rage, or love depending upon the psychodynamics and sociodynamics of the hypnotist-subject relationship and the subject's experiential background. To this is added the typical impulse-handling of the subject intensified by decreased cortical control due to emotional pressure. Emotion is used here to mean a generalized state of visceral reactivity in the sense of Wenger (65) after Lange for whom emotion was vasomotor reaction. It is noteworthy that the hypnotist himself and many observers in an audience may often show the "emotional response" to induction and certainly members of an audience often go into a trance state even though they may not at all be close to the operator. Is this, too, a form of dramatic response? It is hard to believe.

As pointed out by this writer elsewhere (25), often the hypnotist himself becomes overwhelmed by impulses which he must learn to control. What is this? This is the emotional response to human relationship—perhaps a so-called transference, but a state which is biological in its implications and only under control through full resistance to its impact. It is hardly a form of acting, and the hypnotist is certainly not under direct suggestion or probing.

During this state (hypnosis), a reduction of critical thinking is experienced as response to any one or all three of the basic emotions (love, fear, rage). This constriction of critical behavior is not unknown as a response to emotional experience in the everyday situation, varying only with the intensity of the experience and the impulse handling of the subject. Probably this increased emotional state, which is hypnosis, makes it possible for various psychosomatic phenomena to be induced with relative ease in the subject under examination. For example, a subject given the suggestion in hypnosis, that he will be pricked by a pin, will, according to Hull (30), show galvanic reaction. However, as pointed out by the same author, such reaction may be elicited in the waking state also. In this sense the conditioning of emotional response to given stimuli must, of needs, be a factor both in waking and hypnotic suggestion, accentuated in the latter by the fact that the individual is already in a state of behavior marked by lowered threshold for emotional stimuli. Perhaps it is this evidence which has led to a conditioned reflex type of hypnosis theory (52).

In terms of electrical activity, the waking state and hypnosis both seem to give alpha rather than spindle waves. Lindsley (41) points out that during apprehension there may be suppression of the 10 per sec. alpha. After the apprehension is reduced, alpha rhythm may appear with greater voltage, persistence, and regularity. Sensory stimulation likewise suppresses alpha, very much like a startle reaction. In anxiety and prolonged emotional states, there is the presence of either low or high voltage beta, or fast activity of 20-30 per sec. or more, presumably associated with marked affectivity or the lack of it. However, Cohen, quoted by Lindsley, finds in addition to the low amplitude, high frequency beta waves, another type of emotional response showing predominantly high voltage beta waves mixed with alpha waves. Distraction of Cohen's subjects who showed diminished alpha and low amplitude, high frequency beta waves, caused the reappearance of alpha activity. These records, taken from patients, do not help to clarify the relationship between hypnosis and the state of emotion. Perhaps Saul, Davis, and Davis (54) are more suggestive in their study, which shows that suppression of alpha activity is related to activity and excitement as personality characteristics, outwardly manifest, whereas a predominance of alpha appears in the passive, relaxed person. Perhaps passivity and relaxation are the more frequent characteristics of the basic hypnotic state than outward excitement and manic manifestation. This would seem to support a libidinous passivity as underlying at least some aspects of the hypnotic state, that is, assuming that the alpha rhythm dominates in the waking, in the hypnotic, and in the passive, receptive states of behavior (this is, of course, with no reference to anything other than percentage of time that the alpha is present). The present writer believes that were sufficient records available, one would find a general similarity between waking states of emotion and a variety of hypnotic responses in various subjects. It would appear speculatively, that different subjects should respond with different electroencephalographic patterns to the state of hypnotic induction. Gorton (22) refers to the occurrence of E.E.G. changes in age regressions. These may be fundamentally indicators of emotional experience. The area indeed is in need of further investigation.

Reduction in cortical control under the hypnotic state causes the subject to become increasingly susceptible to immediate response of a direct nature to a given stimulus (the response being marked by the subject's ordinary mode of personality handling). In this state, since thinking is inhibited, the control of thought organization and defense is reduced. The subject may blurt out his ideas. However, because of poor thinking integration, the response may be slowed. Such behavior may also occur because of attempts on the part of an emoting subject to introduce rational thinking and to inhibit an action which may seem undesirable. In this way the levels of resistance of the hypnotic state become similar to the modes of handling anger, fear, and love in the waking state. This is the habitual impulse handling based upon previous experience, constitutional factors, or both. Perhaps resistances to hypnotic response may also be accounted for in terms, at least partially, of the process of neural induction, as advanced by Ischlondsky (31) for behavioral states. This may appear as a mild negativism to the state of hypnosis, which, however, the subject cannot fully display (resistance itself may be a type of hypnotic behavior).

Being in a state of emotional disruption, the subject may easily be brought to an extreme expression of the given emotion or emotions involved. This is probably why Bull and Gidro-Frank (20) have been able to make such marked use of hypnotic induction of the emotions. The hypnotic state becomes clearly a means of easier elicitation of various psychosomatic phenomena in the subject. Autonomic representation, both of the para-sympathetic and ortho-sympathetic systems, as indicated by Grinker (23), is present in the cortex. Cortical stimulation of various kinds is certainly involved in emotional response, thus the initial hypnotic stimuli are of an emotional nature, and later act in the feed back fashion (46) to elicit further emotional disruption through cortical stimulation on a kinesthetic-proprioceptive basis. Unquestionable, individual differences in terms of the phenomena that can be elicited in hypnosis, will exist on the basis of differences in the threshold of emotional susceptibility and on the basis of differences in handling impulses that surge within the organisms and are accepted and acted upon, or negated. The similarity to the waking state becomes clearer also when hypnotic behavior is regarded as only another aspect of emotionalized response. It is thus included on a continuum of emotional activity during which human behavior is peculiarly free of what might be regarded as cortical control. Certainly people under profound emotional stimulation may do things against their will. Such behavior may appear compulsive, and the perceptual state of the performer may be distorted, as implied in the sensory-tonic field theory of Werner and Wapner (66). Perhaps in this sense the distortion may also appear in the individual's ability for judgment. Persons in a disturbed emotional state are probably more prone to manipulation by suggestion than are those who are not emotionally upset. It is indeed the thesis of this discussion that hypnosis is a state of emotional reactions, as previously implied. In this sense it fits on the continuum with social movements and psychopathology. Mob behavior is unquestionably much a function of sudden emotional upsurging in an individual, rather than a thought through procedure based upon socio-economic or other reasons. Even Cantril (10) implies that a good deal of the brutality shown in lynchings is the result of mass participation rather than the action of a few perverted individuals. This is emotional contagion, as shown so interestingly in the Orson Welles radio program "The Invasion From Mars" (1938) (11). This was an "unstructured" situation which was expressly built up on emotional disruption. Obviously, persons under such circumstances may easily develop hallucinatory or delusional experiences. A dramatic example of this was the case of the phantom anaesthetist of Mattoon (34), a mass hysteria based on hearsay only, it would seem. Here in Mattoon, Illinois, not only was a story of a prowler developed, but people responded to the "gas" he presumably used, with paralysis, vomiting, etc. This was response to suggestion under emotion. Where these suggestions began is unimportant. Whether they originated in the social milieu, in the ideas of a political leader, or in the thought of an experimenter, given the impetus of emotion, they took root and became reality. During the emotional experience, a wide array of phenomena occur, which have immediate influence on the reasoning and resistance of the subject, and which may also elicit experientially through redintegration, previous response patterns to the given emotional stimuli. Here seems to be the foundation for hypnotic behavior in any and all its forms. Here is a foundation which may be examined in everyday experience, as well as in experimental research.

During such experience as tumescence, for example, persons may become very agreeable and subservient, and do and say many things they otherwise might not. Such behavioral change is known to occur even in the infra human primates. As pointed out by Yerkes (68), when a female is sexually receptive, it is her right to claim control of a situation without interference from her mate. Of course, individual differences in the paired group situation exist. This activity between male and female is a function of the sexual state of the participants at the time. Probably those thinkers, Freud among them, who see hypnotic behavior and love as aspects of the same thing, were influenced by the marked tendency of the subject under each of these circumstances to show inability to resist commands from the dominant party. In sex behavior, as in hypnosis, an emotionally excited person is faced by the problem of impulse handling. Thus sexual acts committed under extreme excitement may cause severe remorse in the participant. In the same sense, an act under anger or fear may cause remorse and uneasiness when the calmed subject becomes aware of it. In fact, behavior under intense emotion may sometimes, and in certain subjects, be performed without an awareness participation on the part of the subject.

The theory of the writer would assume that hypnosis involves always an emotional state, perhaps undefined, or consisting dominantly of one or more of the basic emotions. Such state does not need to be marked by observable excitement, although it may. This emotionality accounts for the heightened suggestibility of the organism, for his "willingness" to give data, for his odd behavior, and for the vast variations in resistance—sometimes with a sudden breaking up of the state. The behavior described is typical of hypnosis and waking suggestibility, but not necessarily so, of sleep.

The person who is emotionally disturbed may accept ideas at variance with reality, depending of course upon his usual relationship to reality. Thus, as pointed out by Burtt (8), there are people who will render confessions of acts they have not committed. Burtt further points out that possibly in colonial times some of the "witches" who were repeatedly told they were in league with the devil, began to believe the statement. Such acceptance of non-reality may of course often be motivated by sheer fear or rage or lust, or any confusional mixture of those that makes the subject uneasy about the "real," and unsure of himself. Probably the confession of Tituba (62) in the Salem witch trials was the result of emotional confusion in a confused and emotionally disrupted environment. The hysteria of the witch trials had the characteristics of a state of mass hypnosis. Here again can be observed the relationship between hypnosis and other emotionally induced phenomena. Probably it was this similarity which led Charcot (14) and Janet (32) to relate hypnosis to hysteria. Janet points out that at least in his cases subjects in unmistakable hypnosis were either suffering from severe hysteria at the time, or else had previously manifested hysterical symptoms. While the present writer would not delimit hypnotic behavior to hysterics, he definitely believes that the handling of the hypnotic state is diagnostically significant (24).

As in the neurotic and psychotic states, a subject in hypnosis is emotionally involved, so that his behavior may be completely or partially out of the voluntary range. In catatonia the resistance to emotional release makes it possible to posture the patient who is so involved in damming the emotions that he cannot shift his posture, and may take postures imposed upon him, as in cerea flexibilitas. This automatic obedience (5), as Kraepelin termed it, may be no more than an inability to reduce resistance to the emotional state, thus the sensorium is overrun, and there is little freedom to respond to the stimuli of the environment selectively.

It is to be noted that the hypnotic subject may in actual practice resist full expression of the emotional state (44), and that this too would add to a reduced cortical vigilance in regard to other responses.

This writer has found, even as Schneck (58), that subjects often equate the hypnotic state with sex behavior, etc. Such subject may show marked resistance or marked coöperation or both. There may be diverse body movement and tumescence of the genitalia. If the subject recognizes the state he is in, there may be outright refusal to submit to hypnotic suggestion. This does not, however, indicate that the subject is not in hypnosis. If there is not refusal, and the subject handles his impulses differently, that is, he cannot resist suggestion, he may later show guilt and perhaps serious emotional disturbance. Such reaction may occur even while the person is in hypnosis, and it would appear quite possible, as reported by Dribben (16), for the hypnotic experience to precipitate a state of pathology. This is particularly so if the conflict remains unresolved.

Regarding hypnosis as a state of heightened emotion places it on a continuum with other states of an emotional nature when the subject is increasingly suggestible, as indicated above. Among these is the state of hysteria and a general reaction of a subject under mob influence. Possibly those people who are more emotionally labile, as hysterics, etc., do make better hypnotic subjects. Playing on the emotions leads to increased suggestibility, and those who are professional propagandists have made much use of this technique in mass movements. Examination of books like *Mein Kampf* (28) for example, reveals tremendous emphasis on emotionally loaded discussion.

This writer has observed in many of his subjects, swallowing reactions, as if to globus hystericus. Some show tremor of the eyelids, and an occasional subject bursts into laughter or tears, perhaps breaking out of the hypnotic state. These seem to be reactions to emotion. In a few cases analysis of the resistance revealed guilt feeling about tumescence as a response to hypnosis. One female subject resisted hypnosis by saying, as if to herself, "No, No." She later remarked about her resistance, "I feel myself sexually stimulated. I think of sex activity, and I refuse to consider it." Indeed it would seem that as a subject adjusts to hypnosis, a given signal initiates the experience of emotion, diffuse or specific, but basic to a heightened suggestibility. The subject at this point is dealing with two types of impulses, those present within and released by the emotion, and those induced by the investigator from without.

It would appear from the evidence, that emotional responses induced hypnotically are genuine. This is the impression derived from the work of Gidro-Frank and Bull (20). Despite some disagreement because the induction of hypnosis itself does not cause wide deviation from the normal in many physiological processes (3), suggestion of emotion under hypnosis has produced very real changes, as demonstrated by Ford and Yeager (17), using the electroencephalograph. It is indeed possible that the mere induction of hypnosis produces emotional response which is of such nature as to be difficult of recording physiologically, unless it be given a specific direction. At such time, a subject's behavior would be more marked by a readiness for full reaction than by full reaction itself; although there are cases where the induction itself produces a marked reaction of distinguishable and definable emotion. The fact however, that hypnotic emotions cause a real change in the physiological picture of response, at least distinguishes hypnosis from play acting, if we are to accept the evidence of Golla (21) who, in studying *PGR*'s of actors, found no galvanic response in the bitterly weeping performer.

If hypnotic behavior on the human level can be related to hypnotic behavior on the animal level, then it appears that the so-called animal hypnosis gives further evidence of the emotional nature of hypnotic response, particularly in view of the fact that often animal hypnosis involves a technique of physical restraint, and it is known from the early research of Watson (64) that physical restraint elicits basic emotional response in the human infant. Völgyesi (63), quoting Kircher in regard to the hypnosis of a fowl, indicates that the placement of the animal's head against the ground and drawing a straight or zig-zag line in front of his beak, induces a condition of immobility. While Marcuse (44) points out that bridging the gap between animal and human hypnosis is difficult, and perhaps misleading, this particular approach seems consistent with parsimony and the general trend of evidence presented in this paper.

What then can be said about the psychophysiologic picture behind the hypnotic state? This must be speculatively presented in the absence of sufficient data. However, in accordance with the foregoing discussion, the following seems to be objectively tenable, pending further demonstration: (*a*) A subject is given continuous suggestions to relax or to sleep. Such suggestions elicit an emotionally loaded response in terms of readiness. This is so both to the suggestion and the operator, if an operator is present. In this respect it can be said that all states involving human relations (directly or indirectly) have emotional impact, and hypnosis is such a state. The impact of such state may be akin to, but not the same as the trophallactic state among lower species of organisms (59). The response given to the suggestion may be intense, accompanied by marked emotion, or it may be too imperceptible to be noticed. Schock (60) has pointed out that less explosive states may in lesser degree produce many of the changes that are observable in anger or fear. These physiological responses he says, are mediated by way of autonomic stimulation or through the action of epinephrin released by the adrenal medulla. It is interesting to notice that injection of epinephrin as carried out by Marañon (45) will elicit responses which are described as if they were like the full emotional states. The experience of Marañon's subjects was reported generally as "cold" emotion. Cantril (12), Cantril and Hunt (13), and Landis and Hunt (40) repeated the work of Marañon, finding generally similar results.

Sometimes the emotion may be completely unrecognized as such, but as maintained by Fremont-Smith (18), the symptoms due to autonomic activity appear in consciousness. It is important to remember that individual differences in somatic manifestations of the emotions exist (60). This has also been demonstrated by Schneck (58) for response of patients to induction.

Pottenger (50) has indicated, as is well recognized, that man's psychic states, so to speak, influence the physical body. If mere talking to John Jones makes him blush or become intestinally disturbed because of emotional stimulation, why cannot hypnosis be considered an extension of this? In hypnosis the person is more vulnerable and less able to make voluntary response, thus he is more emotionally disturbed and more forcibly brought to make suggested compensatory reactions (27).

(*b*) Due to the emotional state there is a decrease in critical control—perhaps inhibition of the cortex. This would be, as is recognized by some investigators, a state that is different from true sleep, or sleep induced by barbiturates, which is regarded as genuine sleep, giving the same brain waves (15). (Ether gives brain waves that differ from those of normal sleep). Pick's (49) work would lead to the conclusion that sleep represents the depression of thalamus and striatum, for these are acted upon by barbiturates (dial, veronal, luminal, etc.). If this is true, then the hypnotic state would involve activation of the subcortical structures and inhibition of the cortex. (It must be remembered that the data here presented has been disputed by Koppanyi (38) who constitutes an equally reliable authority on drug effects.) However, such inhibition would be the result of increased emotional tension and inability to express, because of the lack of relationship between the emotional tension experienced and the reality situation. Under such circumstances, any opportunity to respond is uncritically grasped, and thus response becomes tension reducing, albeit the subject's reaction to his response may reinforce his emotional state because he does not understand his response, or because his response touches upon ordinarily frightening impulse areas. In this way a feed back mechanism (1) is established, and the response may increase the tension and thus may increase the susceptibility to further suggestion: thus most workers have probably found that if the subject can be made to respond to one "extreme" suggestion, others will be more easily responded to. This mechanism may be a learning phenomenon in hypnosis akin to the theoretical position of reinforcement theorists (29). If a situation of the above kind is established, the hypnotic state becomes much like an anxiety state, with obsessive-compulsive aspects, where the anxiety causes the compulsive act which causes further anxiety (although there may be temporary tension reduction), which in turn causes the commission of the act again.

(*c*) How then does hypnosis act therapeutically? On the one hand it may release inhibited emotion. In this way it acts almost in the fashion of the similia similibus curantur principle. Pharmacologically, Cameron (9) has demonstrated such reduction of persistent anxiety states by the introduction of graduated doses of adrenalin.

Furthermore it would seem that an emotional readiness, perhaps even in the sense of anxiety, may be a factor in adjustment to temporary stress (recognizing the detrimental effects of continued stress, as demonstrated by Selye (61)). Thus Jellinek (33) claims that a readiness for electrical shock makes an organism less susceptible to death as a result of the shock. Says Jellinek, about emotional response: "The fact is less well known however, that not only in life, but often also in death, and indeed out of the profoundest depths of the agony, the psychicum quite suddenly, unexpectedly shines forth and is at the same time capable of inspiring the whole organism to some transitory activity, to a functional integration. This strange fact does not even form the basis for scientific discussions dealing not only with the complex question of death, but with the inner comprehension of the psychophysical problem in general." Perhaps the dualism implied here is unnecessary and probably not strictly desired by the writer. It suffices to recognize that emotional stimulation with its concomitant bodily states will affect the organism's responses to such stimuli from the environment as form, color, pain, etc. Hypnosis, if it is a state of diffuse emotion, can understandably be utilized to bring about effects ordinarily elicited through immediate or chronic emotion. Thus psychosomatic disorder, psychopathological states of perception and also rest and relaxation may be furthered by the hypnotic state as well as by the experiences of the individual in interpersonal relations. This thesis places hypnosis on a continuum with other states of emotional excitement, wherein the abient and adient responsiveness of the normal organism may be driven to extremes. Obviously this is a complicated area to examine, and the current theory is frankly speculative.

Summary. This paper presents a theory of hypnotic behavior which attempts to show that the latter is an extension of ordinary emotional responsiveness. The hypnotic state is described as a diffuse state of emotional readiness, unlike sleep, and marked by increased tendencies for abient or adient activity, or compensatory distortions of such activity when it cannot be easily expressed. The interpersonal factor operates here as it does in all person-to-person relations, ranging from parent-child and lover-lover relationships to mob relationships. Physiological and pharmacological evidences supporting this point of view are considered and evaluated.

REFERENCES

1. ASHBY, W. R. The cerebral mechanisms of intelligent behavior. In D. Richter (*Ed.*), *Perspective in Neuropsychiatry*. London: Lewis, 1950. (Pp. 79-95.)
2. BARD, P., & MOUNTCASTLE, V. B. Some forebrain mechanisms involved in the expression of rage with special reference to suppression of angry behavior. *Ass. Res. Nerv. & Ment. Dis., Proc.*, 1948, **27**, 362.
3. BATEMAN, J. F., & PAPEZ, J. W. Significance of the thalamus in psychoses with reference to hallucinations and delusions. *J. Clin. & Exp. Psychopath.*, 1951, **12**, 89-103.
4. BERNHEIM, H. Suggestive therapeutics. C. A. Herter (transl.). New York: London Book Co., 1947 (reissue).
5. BLEULER, E. Textbook of Psychiatry. A. A. Brill (transl.). New York: Dover Publ., 1951 (re-issue).
6. BRAID, J. Neuripnology; or the rationale of nervous sleep, considered in relation to animal magnetism. London: Redway, 1899.

7. Bull, N., & Gidro-Frank, L. Emotions induced and studied in hypnotic subjects: Part II. The findings. *J. Nerv. & Ment. Dis.*, 1950, **112**, 97-120.
8. Burtt, H. E. Legal Psychology. New York: Prentice-Hall, 1931.
9. Cameron, D. E. Adrenalin administration in persistent anxiety states. *Amer. J. Med. Sci.*, 1945, **210**, 281-288.
10. Cantril, H. Psychology of Social Movements. New York: Wiley, 1941.
11. Cantril, H., *et al.* The Invasion from Mars. Princeton: Princeton Univ. Press, 1940.
12. Cantril, H. The rôles of the situation and adrenalin in the induction of emotion. *Amer. J. Psychol.*, 1934, **46**, 568-579.
13. Cantril, H., & Hunt, W. A. Emotional effects produced by the injection of adrenalin. *Amer. J. Psychol.*, 1932, **44**, 300-307.
14. Charcot, J. M. Oeuvres completes: Metallotherapie et hypnotisme. (Tome IX.) Paris: Fourneville & Brissand, 1890.
15. Derbyshire, A. J., Rempel, B., Forbes, A., & Lambert, E. F. The effect of anaesthetics on action potentials in cerebral cortex of the cat. *Amer. J. Physiol.*, 1936, **116**, 577-596.
16. Dribben, I. S. Psychosis following "amateur hypnosis," a case report. *Br. J. Med. Hypnot.*, 1951, **2**, 4-6.
17. Ford, L. F., & Yeager, C. L. Changes in the electroencephalogram in subjects under hypnosis. *Dis. Nerv. Sys.*, 1948, **9**, 190-192.
18. Fremont-Smith, F. The influence of emotional factors upon physiological and pathological processes. *Bull. New York Acad. Med.*, 1939, **15**, 560-569.
19. Freud, S. Group Psychology and the Analysis of the Ego. (Tr. by J. Strachey.) London: Int. Psychoan. Press, 1922.
20. Gidro-Frank, L., & Bull, N. Emotions induced and studied in hypnotic subjects: Part I, the method. *J. Nerv. & Ment. Dis.*, 1950, **3**, 91-100.
21. Golla, F. Croonian lectures. *Lancet,* 1921, **2**, 420.
22. Gorton, B. E. The physiology of hypnosis. *Psychiat. Quart.*, 1949, **23**, 317-343 and 457-485.
23. Grinker, R. R. Neurology. (3rd ed.) Springfield, Ill.: Thomas, 1946.
24. Guze, H. Hypnosis as wish fulfillment: a projective approach. *Br. J. Med. Hypnot.*, 1951, **2**, 6-17.
25. ———. Posthypnotic behavior and personality. *Personal., Sympos. Topcl. Issues,* 1951, **1**, 231-240.
26. ———. The rôle of the hypnotist. *Br. J. Med. Hypnot.*, (in press).
27. ———. Compensatory behavior as completion of the sensory-motor cycle. *J. Genet. Psychol.*, (in press).
28. Hitler, A. Mein Kampf. New York: Reynal & Hitchcock, 1941.
29. Hull, C. L. Principles of Behavior. New York: D. Appleton-Century, 1943.
30. ———. Hypnosis and Suggestibility. New York: D. Appleton Century, 1933.
31. Ischlondsky, N. E. Brain and Behavior. London: Kimpton, 1949.
32. Janet, P. Psychological Healing. (Tr. by Eden & Cedar Paul, 2 vol.) London: Allen & Unwin, 1925.
33. Jellinek, S. Dying, Apparent Death, and Resuscitation. Baltimore: Williams & Wilkins, 1947.
34. Johnson, D. M. The "phantom anesthetist" of Mattoon: A field study of mass hysteria. *J. Abn. & Soc. Psychol.*, 1945, **40**, 187-195.
35. Jones, E. The nature of autosuggestion. *Br. J. Med Psychol.*, 1923, **3**, 206-212.
36. Kline, M. V. Personal Communication, 1951.

37. ———. Hypnosis and diagnostic psychological testing. *Personal., Sympos. Topcl. Issues,* 1951, **1**, 243-252.

38. KOPPANYI, T., & DILL, J. M. Remarks on the distribution of barbiturates in the brain. *J. Pharmacol. & Exp. Ther.,* 1935, **54**, 84-86.

39. KUBIE, L. S., & MARGOLIN, S. The process of hypnotism and the nature of the hypnotic state. *Amer. J. Psychiat.,* 1944, **100**, 611-622.

40. LANDIS, C., & HUNT, W. A. Adrenalin and emotion. *Psychol. Rev.,* 1932, **39**, 467-485.

41. LINDSLEY, D. B. Emotions and the electroencephalogram. In M. L. Reymert (*Ed.*), *Feelings and Emotions.* New York: McGraw-Hill, 1950.

42. LOOMIS, A. L., HARVEY, E. N., & HOBART, G. Brain potentials during hypnosis. *Science,* 1936, **83**, 239-241.

43. MCDOUGALL, W. Outline of Abnormal Psychology. New York: Scribner, 1926.

44. MARCUSE, F. Interpretation in animal hypnosis. *Personal., Sympos. Topcl. Issues,* 1951, **1**, 240-243.

45. MARAÑON, G. Contribution a l'etude de l'action emotive de l'adrenaline. *Rev. Fr. Endocrin.,* 1924, **2**, 301.

46. MEDUNA, L. J. Carbon Dioxide Therapy. Springfield, Ill.: Thomas, 1950.

47. NAUTA, W. J. H. Hypothalamic regulation of sleep in rats; an experimental study. *J. Neurophys.,* 1946, **9**, 285-316.

48. PAVLOV, I. P. Conditioned Reflexes and Psychiatry. (Tr. by W. H. Gantt.) New York: Int. Publ., 1941.

49. PICK, E. P. Pharmacologie des vegetativen nervensystems. *Deut. Zsch. Nervenh.* 1928, **106**, 238-268 and 304-319.

50. POTTENGER, F. M. Symptoms of visceral disease. St. Louis: Mosby, 1944.

51. RIVERS, W. H. R. Instinct and the unconscious. Cambridge: Cambridge Univ. Press, 1922.

52. SALTER, A. What is hypnosis? New York: Smith, 1944.

53. SARBIN, T. Contributions to rôle taking theory: I. Hypnotic behavior. *Psychol. Rev.,* 1950, **57**, 255-270.

54. SAUL, L. J., DAVIS, H., & DAVIS, P. A. Correlations between electroencephalograms and the psychological organization of the individual. *Trans. Amer Neurol. Ass.,* 1937, **63**, 167-169.

55. SCHILDER, P. Quoted in Brenman, M., & Gill, M. M. *Hypnotherapy.* New York: Int. Univ. Press, 1947.

56. SCHILDER, P., & KAUDERS, O. Hypnosis. (Tr. by S. Rothenberg.) *Nerv. Ment. Dis. Monogr. Ser.,* 1927, **46**, 118.

57. SCHNECK, J. M., & KLINE, M. V. Hypnodiagnosis and evaluation of therapy in psychiatry and clinical psychology: Report of a case involving the H.T.P. *Br. J. Med. Hypnot.,* 1951, **2**, 8-19.

58. SCHNECK, J. M. Psychosomatic reactions to the induction of hypnosis. *Dis. Nerv. Sys.,* 1950, **11**.

59. SCHNEIRLA, T. C. Problems in the biopsychology of social organization. *J. Abn. & Soc. Psychol.,* 1946, **41**, 385-401.

60. SCHOCK, N. W. Physiological manifestations of chronic emotional states. In M. L. Reymert (*Ed.*), *Feelings and Emotions.* New York: McGraw-Hill, 1950.

61. SELYE, H. Stress. Montreal: Acta, 1950.

62. STARKEY, M. L. The devil in Massachusetts. New York: Knopf, 1950.

63. VOLGYESI, F. L'Hypnotisme et le mystere cerebral. Paris: Vigot Freres, 1944. (Tr. into French by Andre Mesureur.)

64. WATSON, J. B. Psychology from the Standpoint of a Behaviorist. Philadelphia & London: Lippincott, 1919.

65. WENGER, M. A. Emotion as visceral action: An extension of Lange's theory. In M. L. Reymert (*Ed.*), *Feelings and Emotions*. New York: McGraw-Hill, 1950.

66. WERNER, H., & WAPNER, S. Sensory-tonic field theory of perception. *J. Personal.,* 1949, **18**, 88-107.

67. WHITE, R. W. A preface to the theory of hypnotism. *J. Abn. & Soc. Psychol.,* 1941, **36**, 477-515.

68. YERKES, R. Social behavior of chimpanzees: dominance between mates in relation to sexual status. *J. Comp. Psychol.,* 1940, **30**, 147-186.

Department of Psychology
Long Island University
385 Flatbush Avenue Ext.
Brooklyn 1, New York

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 329-351.

PROBLEM-SOLVING BEHAVIOR IN MEN WITH FRONTAL OR OCCIPITAL BRAIN INJURIES*[1]

Department of Neurology and Psychiatry, New York University College of Medicine

W. S. BATTERSBY,[2] H. L. TEUBER, AND M. B. BENDER

A. THE PROBLEM

In previous communications (56, 57, 58) we have described performance of complex visual discriminations in a group of men with battle wounds of the brain and in a control group of veterans with peripheral nerve injuries. These studies demonstrated that lasting deficits in the recognition of hidden figures (56), and in solution of certain sorting tests (1, 58) can be produced by penetrating cerebral lesions in man. The defects in performance varied in severity and appeared in some cases of brain injury but not in all; a considerable number of brain injured subjects equalled or excelled the controls on individual tasks. Moreover, men with massive frontal penetrations did not do any more poorly than did men with similarly severe parieto-occipital penetrations. In general, there was no indication that lesions in the frontal lobes are necessarily more disabling in performance on complex visual discriminations than lesions in the posterior lobe substance. These findings seem to contradict the usual assumption of a predominant importance of the frontal lobes for intellectual functioning in man (4, 13, 17, 23, 24, 51, 52). Our findings are in agreement, however, with results recently reported for subhuman primates (2, 3, 6, 26), and other studies in man where negative results have been obtained after frontal lesions (28, 29, 30, 36, 47).

The question must be raised, however, whether the complex visual tasks employed in these previous studies were appropriate indicators of intellectual functions. Conceivably, tasks which do not directly depend upon visual sensation for their solution might reveal intellectual deficits more specifically related to the symptoms of frontal lesions. In addition, it is desirable to employ tests which are more life-like than the tasks used thus far. In the

---

*Received in the Editorial Office on December 6, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] From the Psychophysiological Laboratory, Department of Neurology and Psychiatry, New York University College of Medicine. This work was carried out (as part of a larger project on the After-Effects of Brain Injuries of World War II) under Contract No. V1001M-176 with the United States Veterans Administration.

[2] Present Address: Department of Neurology, The Mount Sinai Hospital, New York City.

present study we decided therefore to introduce a task involving "reasoning" in an everyday setting. The use of such problems has been advocated in experimental investigations of reasoning in normal subhuman primates (31, 34, 35) and man (7, 8, 20, 62, 63). Moreover, certain advantages of such an approach over the use of puzzles (50), paper-and-pencil tasks (5, 39), and various tests of concept formation (13, 35, 61) have been demonstrated by N. R. F. Maier (40, 41, 42). One of Maier's more familiar tasks, the two-string problem (41), was adapted for the present experiments.

The principal questions we hoped to answer by means of this task were the following: (*a*) To what extent does brain injury in man affect problem solving in a Maier-type reasoning task? (*b*) If problem-solving is impaired by brain injury, are frontal injuries more disabling than comparable injuries in the posterior (parieto-occipital) lobe substance? (*c*) What qualitative changes in problem solving behavior, if any, can be noted in the injured group?

B. Subjects Studied

The subjects of this experiment were drawn from the same general population of brain injured men who have been described in previous reports from this laboratory (56, 57). A total of 80 subjects were studied, all veterans of World War II. As in the previous studies, there were a number of subgroups: Group *A* (anterior lesions) consisted of 20 men with penetrating gunshot wounds of the frontal lobes; there was no indication of visual field defects in these patients. Group *P* (posterior lesions) consisted of 20 men with similar wounds of either the posterior parietal or occipital lobes or both (parieto-occipital lesions); all subjects in this group had homonymous field defects. Loss of brain substance in Groups *A* and *P* had been established prior to this study by the methods employed previously (57). Group *C* (controls) consisted of 40 men with peripheral nerve injuries sustained in combat in World War II. None of the controls had visual field defects, nor did they have signs or history of head injury. Groups *A* and *P*, and half of *C* participated in Experiment I. The remaining 20 cases in Group *C* were employed in a control experiment (Experiment II).

C. Experiment I: Solution of the String Problem by Brain Injured Subjects and Normal Controls

1. *Procedure*

A modification of Maier's string problem (41) was used. The subject (*S*) entered a room in which two strings were hanging freely from the ceiling. His task was to tie the ends of the two strings together although they were too far apart to permit holding one while grasping the other. Each *S* was told not only that the problem could be solved, but that it had to be solved without cutting either of the strings. Moreover, he was informed that the only objects he could use were those which would be provided by the experimenter (*E*). The "correct" solution was to tie any of these objects to one of the strings, converting the string into a pendulum, and swinging it toward the other string.

All *S*'s were tested individually. At the beginning of the testing session, *E* and *S* sat at a desk at one end of the room. After the instruction had been given as described, and the position of the strings pointed out, *E* started a stopwatch and placed a pair of scissors in plain view of *S* on top of the desk. At successive two-minute intervals, *E* added the following objects, one at a time, to the scissors on the desk: clothes pin, small pulley, child's yo-yo, and fish-line sinker. These objects were intended to serve as a graded series of hints of increasing relevance to the problem situation. If *S* had not solved the problem at the end of 10 minutes, he was given a series of five verbal hints until solution was achieved. In sequence, these hints indicated that: (*a*) any *one* of the objects could be used to solve the problem, (*b*) the objects did not *have* to be used in their customary way, (*c*) *all* the objects had one common property which was all that was needed, (*d*) all the objects had *weight*, and (*e*) if one of the strings would come to the center, the problem would be easy.

In all cases the time required for solution was recorded and a detailed protocol of *S*'s behavior kept. Additional test data on a modified Weigl-type sorting test (21, 61) and scores on a paper-and-pencil version of Gottschaldt's "hidden figures" (18, 19) were available for most of the subjects from previous studies (57).

2. *Results*

a. Quantitative data. As can be seen from the frequency distributions in Figure 1, the time required for solving the string problem varied greatly within all groups. In any one group, some of the subjects solved the problem quite rapidly, whereas other subjects took a comparatively long time (regardless of whether they had cerebral lesions or not). There were also, in each group, several subjects who took a rather long time to discover the pendulum principle, and, hence, required a greater number of practical "hints" toward solution. The frequencies of solution, per group, are itemized in Table 1 according to the object which served as a "hint" to solution. The results suggest that prolonged solution periods (exceeding two "hints") were rare in Group *C* (controls), and more frequent in Groups *A* (frontal lesions),

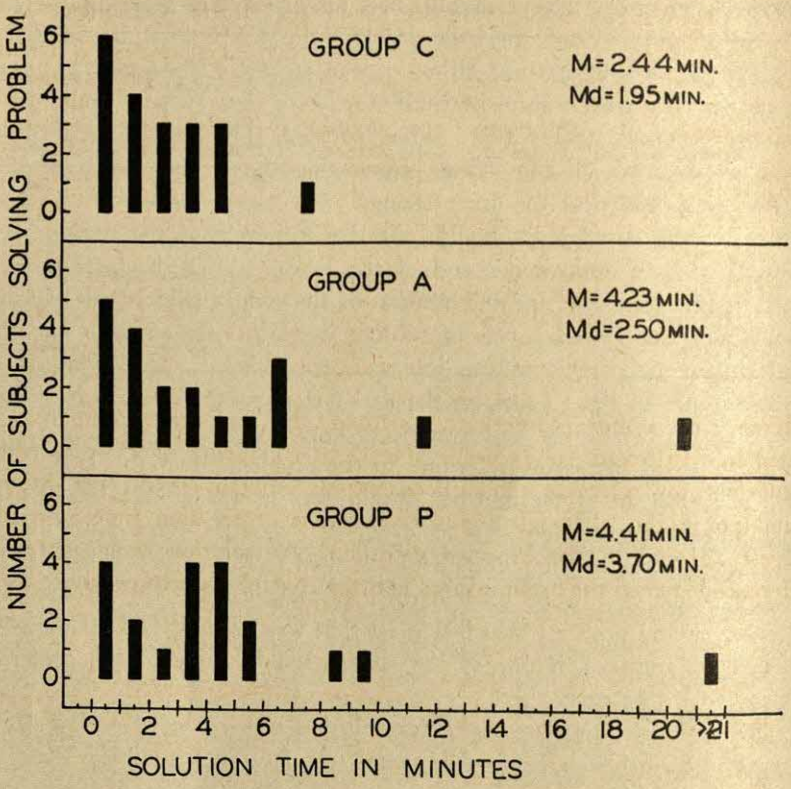


FIGURE 1
Frequency distributions of solution time for groups in Experiment I; *C* (controls), *A* (frontal lesions), and *P* (parieto-occipital lesions). Mean and median time for each group given on right.

and *P* (parieto-occipital lesions). The difference between Groups *A* and *P* is small; that between either Group *A* or *P* and the control Group, *C*, is more conspicuous.

Evaluation of these frequencies by the chi square technique shows that the difference between *A* and *P* (anterior vs. posterior lesions) is not significant. The difference between either *A* or *P* (or both), on the one hand, and *C* (the control group), on the other, is significant at less than the one per cent level, if one accepts the frequencies of the control group as representative of the population. However, this test of significance is based on one cell frequency of two (for the control group), a condition which precludes drawing valid conclusions from the chi square test (38). Evaluation of the same

TABLE 1
FREQUENCIES OF SOLUTION PER GROUP FOR "HINTS" PRESENTED

		Frequencies Per Group		
Solution time	"Hints" given	Control subjects	Patients with frontal lesions	Patients with parieto-occipital lesions
0.1-2.0	scissors	11	9	6
2.1-4.0	scissors, clothespin	7	4	5
> 4.1	one or more of following: pulley, yo-yo, sinker, verbal hints	2	7	9
Totals		20	20	20

difference by a more conservative method, the rank-order comparison suggested by Wilcoxon (64), yields a level of confidence of $P = .8$, which is insufficient for rejecting the null hypothesis. In any event, the differences found are slight, although cumulative graphs of solution time (Figure 2) suggest a trend in the expected direction. As solution times increase the difference between the brain injured and the control *S*'s becomes more marked.

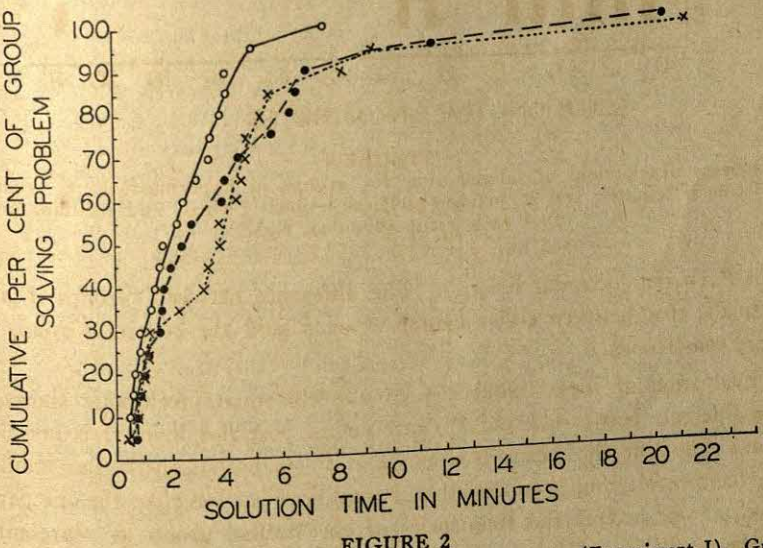


FIGURE 2
Cumulative percentage plot of solution time for each group (Experiment I). Group *C* (controls)—open circles; Group *A* (frontal lesions)—solid circles; Group *P* (parieto-occipital lesions)—crosses.

In summary, many of the brain injured *S*'s did as well on the string problem as did the controls. However, seven men in the group with anterior lesions and nine in the group with posterior lesions took more than four minutes. In the control group, all but two solved the problem in less than four minutes. It is of special interest therefore to see whether some of the very "slow" problem solvers (in Groups *A* and *P*, or *C*) proceeded in a manner qualitatively different from that of the "fast" subjects.

b. *Qualitative data.*

(1). *"Fast" controls* (*using scissors or clothes pin*). The manipulative behavior and verbal comments of control *S*'s, during and after solution, were reasonably uniform for those who solved the problem in less than four minutes (18 of 20 controls). On presentation of the scissors, these "fast" subjects hesitated briefly, usually stating that there must be a "catch" or "trick" since the scissors could not be used in the customary way. (*E* had explicitly ruled out any cutting of the strings.) Soon afterwards, and not later than following presentation of the clothes pin, these subjects would comment that they could "see" the connection: they had to use the presented object (scissors or clothes pin) in a "new" fashion, namely as a pendulum bob, in order to swing one string towards the other. Illustrative verbatim comments obtained during and after solution follow:

> *Control 8:* Looks at scissors, lifts them, says "these are no good, if I can't cut them (viz., the strings). Just a confusion device." Puts scissors back on desk, looks around, tries to use his chair (apparently intending to tie one string to the back of the chair). Is reminded by *E:* "use only the objects I give you." Stands still, then goes quickly back to desk, takes scissors, ties them to one string, swings it towards the other. Time to solution: 52 seconds. Asked by *E* how he had solved the problem, *S* says: "I couldn't reach one (string) from the other, and I couldn't use the chair because you stopped me. Then I got the idea of swinging it (string) over and thought of using the scissors as a weight on a pendulum. You told me in the beginning I couldn't cut the strings, so I had to use the scissors in a different way."
>
> Similar retrospective accounts were given by other "fast" subjects, e.g., Control 6: "Scissors themselves (were) no good if I couldn't cut the strings. Must have some other connection—the only thing was weight. I couldn't reach one string over to the other, but I could swing it over." (Time to solution: 27 seconds.)
>
> *Control 19:* "I realized near the beginning I had to make one string come to me, but didn't think the scissors would help because you said I couldn't cut the strings. Then suddenly I saw (sic) that the scissors could be used as a weight; it was the only thing available." Time to solution: 1 minute, 25 seconds.

Control 20: Tries to reach one string while holding the other, is unsuccessful but keeps at it for over a minute. Picks up scissors, puts it down, says "that is no good." Keeps trying to reach one string from another, still unsuccessful. *E* presents clothes pin. *S* looks at it, then quickly picks up *scissors,* ties it to one string and swings this string towards the other. Solution time 2 minutes, 23 seconds. *S*'s retrospective comment: "I didn't see (sic) the connection between the scissors and the string until you gave me the clothes pin. The clothes pin was easier to see as a weight, for some reason, and then I realized I could swing the string over. I used the scissors because it was even heavier (than the clothes pin) and would work better."

Control 5: "I couldn't reach it (the string) so it had to come to me. Thought of a weight and swinging it over. The clothes pin gave it away." (Time to solution: 2 minutes, 52 seconds.)

Control 4: "Didn't think of the scissors as a weight. The clothes pin looked better." (Time to solution: 3 minutes, 20 seconds.)

(2). *"Slow" control subjects (exceeding four minutes and requiring more than two hints)*. Of special interest are behavior and comments of the two control subjects who required more than four minutes to solve the problem.

Control 15: At start of the testing session, *S* tries to reach one string while holding the other—says "can't reach it." Takes scissors, opens them, ties one ring to one string, "points" other end of scissors toward the other string. Says, "this is hopeless, I can't get them together." Keeps trying to reach the other string. (Clothes pin given.) Looks at clothes pin, says "that's no good," puts it back on desk, continues attempts to reach other string. (Pulley given.) Looks at pulley, says "I should have done this right at the beginning," ties pulley to one string and swings it toward the other. At end of test, *S* says: "It wasn't until I saw the pulley I realized I could swing the string. The little loop on top of the pulley (viz., the ring for attaching pulley) gave me the idea, I think. Afterwards I thought, God, I could have done that same thing with the scissors." Time to solution: 4 minutes, 57 seconds.

Control 18: At beginning of testing, *S* looks at scissors and asks "But can't cut them?" *E*, "No cutting!" *S* tries unsuccessfully to reach one string while holding the other, says "It's like a trick, don't see how to get it." (Clothes pin given.) *S* says, "I can't stretch them to get them together." Studies situation, measures strings towards center of room to see if they are long enough to reach each other. *S* comments during test, repeatedly, "can't cut them . . . don't see how to do it, these strings are too short. . . . A clothes pin!—now, what the hell? A pulley! how will that help me? None of these objects will help." *E* puts yo-yo on desk. *S* picks it up and says: "Can I use the scissors only to cut?" *E:* "You can use the scissors any way you wish." *S* solves problem with scissors using the pendulum method. In retrospect *S* says: "I kept thinking that the scissors would only cut or extend one string a little. Then I saw the yo-yo, and the idea of the scissors as a weight for a pendulum hit me." Time to solution: 7 minutes, 31 seconds.

(3). *"Fast" brain injured subjects (using scissors or clothes pin).* Behavior and comments of the brain injured subjects who solved the string problem on the basis of the initial hints (scissors or clothes pin) were quite similar to those of the "fast" control subjects, although the brain injured subjects took slightly longer than the controls. A protocol of one of these "fast" patients is presented in illustration.

> *Case A19:* Looks at scissors, says: "These scissors, get them out of here! I can't cut or break them (strings)." Tries to reach one string while holding the other—no success. Studies problem for a moment, looks at scissors, then suddenly ties them onto one string and swings it toward the other. Solution time 1 minute, 55 seconds. Immediately after solution *S* comments: "At first, I didn't see the scissors as any help. Then I realized I had to use them in a different way, if it was to help any; then I thought of a weight and of swinging. It was the only way the scissors would help. It had to be right." It is interesting to note that this man had massive bilateral frontal lobe damage and was considered by most observers as unduly facetious, euphoric, given to exaggerations and tactless remarks. Nevertheless, he solved the string problem in less than two minutes. Three months later, during another testing situation, the patient reminisced: "Giving the scissors was a bad steer, but it was the only thing that could be used, so it was easy to figure out it had to be used in a different way."

(4). *"Slow" brain injured subjects (exceeding four minutes and requiring more than two hints).* The comments of the brain injured *S*'s who required more than the initial two hints suggest that they had difficulty in "seeing" the usefulness of either the scissors or clothes pin as a weight on a pendulum. There were seven "slow" subjects in Group *A* (20 cases with frontal lesions), and nine in Group *P* (20 cases with parieto-occipital damage). The following protocols are characteristic of their problem solving behavior:

> *Case P6:* Tries to reach one string while holding the other—no success. Now opens scissors and tries to use it to make one string slightly longer—still no success. Asks, "can't cut it? Can I use a chair?" (*E* says, "use only objects given.") *S* seems perplexed, says, "Impossible." (Clothes pin given.) *S* looks at clothes pin, says, "Can't see how to tie them, I can't reach it." (Pulley given.) Picks up pulley, looks at string, suddenly ties pulley on string and swings it toward the other. Solution time 5 minutes, 10 seconds. In retrospect, this patient said: "When I saw the pulley I realized right away I could use it as a weight and swing the string. I think the hook on top of it (pulley) gave me the idea."

Case P2: Tries to reach one string while holding the other—no success. Then suddenly raises one leg (the patient was unusually tall) and pulls in the second string, while holding the first. (Is told by *E* that instructions limit him to the use of his hands and of the objects given by *E*.) *S* objects: "But where do I get some more string to do it with?" Seems perplexed. (Clothes pin given.) Picks up clothes pin, says, "How do they (strings) get together?" Keeps trying to reach the second string, says "This is something (the problem)—must remember this!" (Pulley given.) *S* looks at pulley, his face lights up, and he says, "Maybe I could swing it over." Solves problem using pulley. Solution time 4 minutes, 40 seconds. Retrospective comment: "Being heavy, the weight of the pulley gave me the idea. The scissors and clothes pin confused me. I thought of the scissors as cutting strings, but you told me I couldn't do that, and the clothes pin didn't seem to have any weight. When you gave me the pulley I saw I could use it as a weight."

Most revealing were the comments of those brain injured subjects who required many specific hints and took a relatively long time for solving the problem. All of these exceptionally slow subjects needed the hint given by the fish-line sinker; many needed the additional verbal hints given at the end of 10 minutes. Typical protocols are the following:

Case P13: *S* tries to reach one string while holding the other—no success; says "How to get them together?" Inspects situation, looks baffled. (Clothes pin given.) "Say, a clothes pin!" Picks it up, puts it back on desk, says "All right, I must look for the twist. I'm stymied." Tries to tie one string on cuff button of coat is prevented by *E* from going further. Says "I can't cut or break the strings!" (Pulley given.) Picks it up, says "This thing's no good. You know I even thought of throwing one string up in the air and trying to catch it after I got the other. (!) If I only could keep one here in the middle. I just can't suspend it in mid-air. (Yo-yo given.) "A yo-yo—what's the idea?" Wraps one string around yo-yo, says "Thought I could use it as a pulley." (Fish-line sinker given.) Looks at it says "weight!" Suddenly smiles: "I've just got my first idea which I think will do it. It probably would have worked with any of the others." Ties sinker on string and solves problem. (Solution time: 9 minutes, 14 seconds.) At end of session *S* says "I knew right away I had to get the string to come to me, but it wasn't until I saw the weight (sinker) that the pendulum idea came to me."

Case A12: At beginning, asks: "Can't cut them?" Tries to reach the second string while holding the first. Says "How can I get it over there (pointing)." Studies problem for a minute, ignores scissors, says "It can't be done." (Is reassured by *E* that problem can be solved.) Looks at clothes pin, says "That's out of the question." Looks around the room, sees a meter rule, says "Can I use the ruler?" (*E* repeats instructions "Use only the objects I give you.") Now *S* looks at ceiling connections, says "Maybe I can pull one down." Looks at pulley, ignores it, takes a chair, says, "I can't use this either?" (no) Tries to use scissors, opened widely, as an "extension" of the string, says "No good, too short." Looks at yo-yo, says "There is no string on it, so that's out. Plays with objects, but seems helpless, says "It's impossible." Looks at fish-line sinker: "A weight, whichever object I used wouldn't help me. I can only do it if I have a big object to reach one from the other." Tries using several of the objects to extend one string and fails. Indicates he is ready to give up. (Is given verbal hint No. 1 by *E:* "Any one of the objects could be used to solve the problem.") "I just got it, it just dawned on me." Takes sinker and solves problem; also uses pulley, tying it to the other string, swinging both! Then says "The fish line weight gave me the idea—the others didn't help. Seeing the weight as a weight (!) gave me the idea." (Solution time: 11 minutes, 27 seconds.)

The extreme case is represented by a patient in Group *P*, Case 1, whose solution time was 21 minutes and 15 seconds.

Case P1: At beginning ignores scissors, says: "I need a piece of string or wire in order to reach the other one." Tries to reach one string, holding the other. Looks at clothes pin: "Still need string, clothes pin won't help." Seems unable to go on, glances at pulley—"can't be done." "Scissors could be used to cut a piece of string from someplace else, but where?" Continues trying to reach from one string to the other. (Yo-yo given.) "A yo-yo—that's not going to help. I need about 4 feet of string." (Sinker) "Weight's no good, I need a piece of string. I don't see (sic) how any of these things will help." (Hint 1.) Picks up pulley, ties it on one string, tries to use it as an extension device—"No, can't be done, need more string." (Uses fish line sinker in same way with same results.) (Hint 2.) Studies objects more carefully, but still can't get started. (Hint 3.) —Goes through objects one by one, naming their customary function, then says "For each one you need string and I don't have any." (Hint 4.) Studies problem at great length, says: "I still need string," but seems to be less sure. (Hint 5.) Says immediately, "Yes, it would help if it would come to me, but I still need more string." Is highly motivated, however, considers strings from all angles, suddenly smiles, ties fish line sinker on one, swings it and solves the problem. Solution time: 21 minutes, 15 seconds. At conclusion says "O.K., anyone of them could have done it, but the fish line weight looked more like a weight to me. (!) It gave me the idea; the other objects didn't have any direct use. I thought of a surveyor's plumb bob and then I had it!"

3. *Comment* (*Experiment I*)

Our quantitative results fail to show any obvious (significant) difference between groups of subjects. Nevertheless, there are a few patients with brain injury who required a conspicuously longer period to solve the problem, and throughout the groups there is a trend towards longer solution times among the brain injured *S*s. Men with frontal injuries (Group *A*) did not do any worse, quantitatively or qualitatively, than those with parieto-occipital lesions (Group *P*). None of the group differences in performance noted were of the magnitude one would expect on the basis of the earlier literature, where frontal lobe lesions were claimed to produce specific deficits in "intelligence" (12, 14, 23, 24), "reasoning" (24, 53, 55), "imagination" (10, 11), "planned administration" (48), "synthesis" (4), "maintenance of set in the face of interference" (45, 46), and "loss of the abstract attitude" (15, 16, 17).

Admitting that specific "frontal lobe deficits" failed to appear on comparing our groups, we might still ask whether the few extremely slow patients (regardless of site of lesion) performed in a manner qualitatively different from their controls. All of the protocols suggested that the hints given by *E* were instrumental in guiding the *S*s toward a solution of the problem. The most general comment made by the slower subjects was "I couldn't see that I had to use a weight until you showed me the pulley (yo-yo or sinker)." A dependence on what seemed to be more specific hints was apparent in *some* of the brain injured subjects, but it did not occur in a sufficiently large number to permit one to generalize about the effects of brain injury on problem solving.

D. EXPERIMENT II: EFFECT OF INSTRUCTION ON THE SOLUTION OF THE STRING PROBLEM BY NORMAL SUBJECTS

Our interpretations of results of Experiment I presuppose that the five objects employed as "practical hints" formed, in fact, a graded series of increasingly direct aids toward solution. Experiment II was undertaken to further define the rôle of these hints.

1. *Procedure*

In the preceding experiment, solutions of the string problem were obtained under instructions designed to restrict the *S*s to the specific objects presented by *E* ("Use only the objects I give you from time to time"). The second experiment was undertaken to explore the rôle of these "restricting" instructions, by contrasting them with "less restricted" and "unrestricted" situations. The 20 remaining *S*s in Group *C* (normal control *S*s) were employed for this purpose. Ten of these men were given the string problem under "less restricting" instructions than had been given in Experiment I, and the remaining 10 were tested under essentially "unrestricted" conditions.

a. Less restricted instructions. The 10 *S*s who received the less restricted instructions were told that they could solve the string problem by using any object in the experimental room that they might want to use, including the objects which *E* would put before them on the desk, at two-minute intervals. These *S*s, however, were told that they would have to find and use the particular method of solution which *E* had in mind. *E* observed *S*'s presolution behavior, and at two-minute intervals supplied the same five objects (practical hints) in the same order, as in Experiment I. As soon as *S* began to use any method other than that of converting one of the strings into a pendulum, he was told that he was "wrong," and encouraged to try a different approach.

b. Unrestricted instructions. The remaining 10 *S*s, the unrestricted group, worked under the same conditions as all others, except that the five objects (scissors, clothes pin, pulley, yo-yo, sinker) were haphazardly placed on a table before they entered the experimental room. This table was at one end of the room opposite to the desk, at which *E* and *S* were sitting at the beginning of the trial period. *S*'s attention was not drawn to these five objects at the opposite end of the room until 20 minutes had elapsed. If, by that time, *S* had not solved the problem by the correct (pendulum) method, he was given verbal hint No. 1 (indicating that the objects on the table had something to do with the problem). At succeeding two-minute intervals he was given the additional verbal hints, in the same form and in the same order as in Experiment I, until he arrived at the solution.

2. *Results*

a. Quantitative data: Time to solution. The frequency distributions of solution times in Figure 3 indicate that the string problem was made more difficult when *S* was not restricted by instruction to the use of the five stimulus objects. Under less restricted instruction, the performance of the 10 control subjects was markedly poorer than that of the 20 controls used in Experiment I (restricted instructions). The group of 10 controls who worked under "unrestricted" instruction, i.e., whose attention was not drawn from the start to the five objects, did poorest of all. To be sure, there is some overlap between groups, as can be seen from the cumulative graphs given in Figure 4. These cumulative graphs illustrate how changing instructions, from restricted to less restricted, and finally to unrestricted, led to increasing prolongation of the time necessary for solving the string problem.

b. Quantitative data: Choice of objects. In Experiment I and (partly) II, the five objects which served as practical hints were presented in a fixed order (from scissors to fish-line sinker). This order was assumed to represent a graded series of increasingly direct hints towards solution. Under unrestricted instructions, in Experiment II, a "free choice" situation was used; that is,

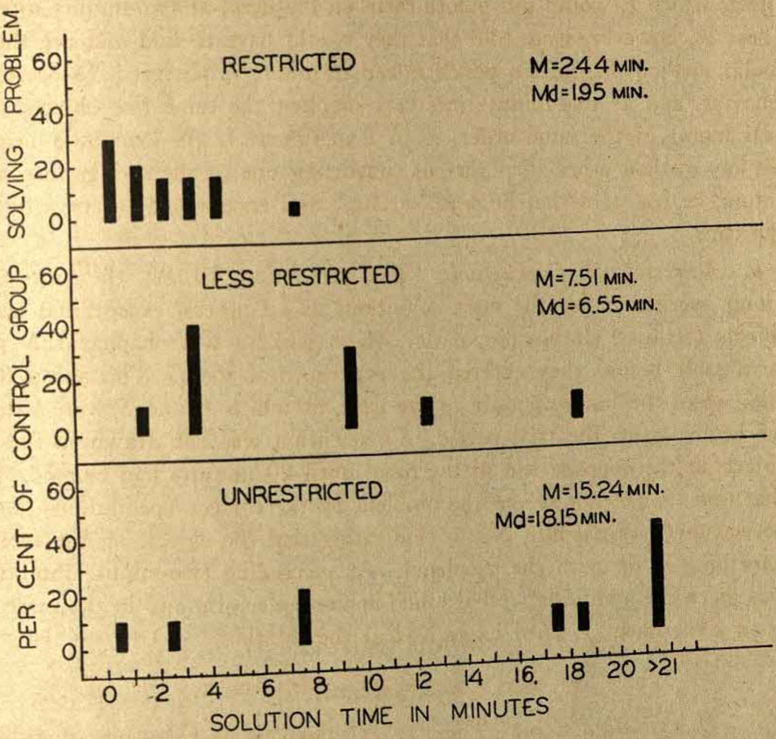


FIGURE 3
Frequency distributions (in percentage form) of solution time for three control groups (Experiment II), working under restricted, less restricted, and unrestricted instructions. Mean and median time for each group given on right.

all five objects were available simultaneously from the start. Table 2 shows the frequency with which each object was used in Experiment II for solution of the problem. Of the 10 subjects in the unrestricted group, six used the sinker, three the yo-yo, one the pulley, in order to convert one string into a pendulum and thus solve the problem. None of the subjects used scissors or clothes pin. We may conclude, therefore, that the five objects did form, approximately, a graded series of hints with increasing "obviousness," as one progressed from scissors to sinker.

c. *Qualitative data.* The protocols show that our normal *Ss* differed greatly in approach to the problem, as the instructions were varied. This was in contrast to the behavior of the normal controls in Experiment I (uniform, restricted instructions), in which all but two *Ss* (of 20 normals) solved the

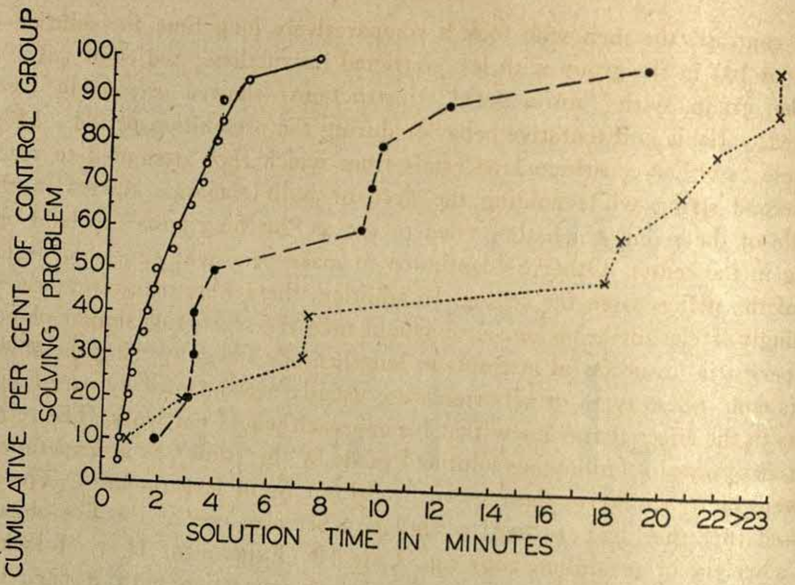


FIGURE 4
Cumulative percentage plot of solution time for each control group (Experiment II). Group with restricted instructions—open circles; Less restricted instructions—solid circles; Unrestricted instructions—crosses.

TABLE 2
FREQUENCIES OF SOLUTION PER "HINT" (OBJECT)—EXP. 2

Object	Restricted	Type of instruction Less restricted	Unrestricted	All types
Scissors	11	2	0	13
Clothespin	7	4	0	11
Pulley	1	0	1	2
Yo-Yo	1	1	3	5
Sinker	0	3	6	9
Totals	20	10	10	40

problem quickly, and commented in various terms (but with identical meaning), that the crucial step towards solution had been to "see" a new or different use for the scissors or clothes pin. Under "less restricted" instructions (10 normal controls), the six *S*s who solved the problem rapidly, commented (as did the "fast" subjects in Experiment I) on the fact that the scissors or clothes pin had to be used in a "tricky" (not customary) way. These men ignored the other objects in the experimental room, although a number of objects in the room could have served as weights.

By contrast, the men who took a comparatively long time for solution—four (of 10) in the group with less restricted instructions; and eight (of 10) in the group with "unrestricted" instructions—showed extremely complicated, variable and tentative behavior during the presolution period. Most of these *S*s either constructed pedestals from which they attempted to reach the second string while holding the first, or built elaborate towers in the middle of the room which they tried to use as "hitching posts" to hold one string in the center. Other *S*s continued to make persistent efforts to detach one of the strings from the ceiling. In addition, there were many instances of seemingly irrelevant behavior: an *S* would take the scissors or similar object, and persist in unsuccessful attempts to lengthen one string by tying the object on its end. These types of behavior were usually accompanied by overt comments to the effect that *S* knew that his approach would not work. The retrospective comments (following solution) made by the "slow" *S*s in Experiment II were quite similar to those given by "slow" *S*s in Experiment I. The *S*s stressed that they had to "see" a similarity between one of the five objects and a weight or pendulum bob; e.g., Case 10 (Experiment II): "I looked at the yo-yo and figured it had to have some connection with it (i.e., the solution of the problem). None of my other attempts was any good. Then I realized the yo-yo could be used to swing one string. I guess the idea of a yo-yo usually being at the end of a string helped me."

Another *S*, after many unsuccessful attempts at solution, said "I got it now" —looking at the fish-line sinker. Picks up sinker, ties it to one string, swings string over to other string, and ties both strings together. Then: "I looked around the room and saw the sinker, and in one flash I realized I could swing it, none of the other ways was right."

3. *Comment* (*Experiment II*)

Results from Experiment II supplement those from Experiment I. Under the "less restricted" and "unrestricted" instruction of Experiment II, normal controls, as a group, took much longer to solve the string problem than did either the normal or brain injured *S*s who worked under the "restricted instuctions" of Experiment I. One may conclude therefore that an extension of the range of possible tools for solution tends to retard solution.

When the *S*s had a free choice of objects, that is when scissors, clothes pin, pulley, yo-yo and sinker were all simultaneously available from the beginning of the test (unrestricted instructions), they chose only the last three objects, most often the sinker—and did not use either scissors or clothes pin. The relative ease with which any of our stimulus objects could be recognized as an appropriate tool varied roughly in the order assumed to exist in Experiment I. Apparently, it is "easy to see," as our *S*s said, that a fish-line sinker may serve as a weight, but it is much less easy to see that a pair of scissors or a clothes pin could be used in the same fashion.

One could describe the difficulty of our slow *S*s, under all conditions of testing, as an inability to disregard the customary use of scissors or clothes pin. This formulation might have been suggested by several theories of "reasoning" (8, 32, 40, 41, 42, 50, 54, 62, 63) which stress that the main obstacle to successful problem solving is "functional fixedness" (8, 62, 63) or, the inability to relinquish "obvious" but irrelevant approaches to solution (8, 20, 41, 42). In most of these theories the presolution period is set in parallel with the early stages of inspection of a puzzle picture (18, 19): the stage during which the concealed figure remains undetected. The process by which the concealed figure is suddenly perceived, usually labeled "restructuring," or "reorganization," is considered in these theories as identical with the process leading to the ultimate solution in a practical task such as the Maier string problem.

Such an identity of process implies correlation, and Maier has indeed postulated that ability to solve problems of the sort discussed should be positively correlated with ability to detect hidden figures in puzzle pictures (18, 19, 44). We have made a preliminary search for such correlations by comparing the performance of our subjects on the modified string problem with that on several of the complex visual discriminations on which they had previously been tested (57).

4. *Relation of Ability to Solve the String Problem to Ability for Making Complex Visual Discriminations.*

In line with Maier's prediction, a significant positive correlation was found between performance on the string problems (under "restricted instruction") and on Thurstone's modification of Gottschaldt's figures (59). However, this correlation obtained only for our control group. As Table 3

TABLE 3
CORRELATION COEFFICIENTS BETWEEN TESTS FOR EACH GROUP

Groups	Tests			
	String-problem: Gottschaldt		String-problem: Sorting	
	r	*P*	*r*	*P*
C (Controls; N = 20)	+0.72	.01	+0.46	.05
A (Frontal lobe lesions; N = 19)	+0.22	.40	−0.01	.90
P (Parieto-occipital lesions; N = 18)	+0.21	.50	+0.13	.60

shows, no significant correlation between performance on the string problem and on the Gottschaldt figures could be demonstrated for either of the brain injured groups (Groups *A* and *P*).

The correlations between performance on a modified sorting problem similar to Weigl's test (21, 61) turned out to be low for the control group (although significant), and insignificant for the brain injured groups (see Table 3).

Apparently the ability of the normals to perform well on the string problem is more closely paralleled by ability to unravel concealed figures than by ability on a task involving "categorizing" or sorting to criterion. This finding recalls the curious insistence of our subjects that they "saw" the solution or "couldn't see" how the scissors might serve as a weight. The absence of any significant inter-test correlation in the brain injured groups precludes any simple interpretation; the causes of failure among the slow subjects may have been heterogeneous, despite the similarity of their overt behavior.[3]

E. General Discussion

It is evident from the main experiment reported here (Experiment I) that some deficits in problem solving may appear following brain injury in the frontal or parieto-occipital areas. However, with the task in question, these deficits are not particularly marked; certainly not any more marked for frontal than for parieto-occipital lesions. This outcome is quite similar to that previously reported from this laboratory (56, 57), where a large number of tasks requiring complex visual discriminations yielded the same results. There are deficits following frontal lesions, but equal or greater deficits following lesions of the posterior lobe substance. Furthermore, the present report agrees with the earlier studies with regard to the wide range of abilities found among the brain injured patients. On the string problem, as well as on the visual tasks, some patients always equalled or surpassed the average normal control. The difference in favor of the controls, on all tests employed, thus far, in these studies, are differences between overlapping groups.

Similarly instructive is the outcome of Experiment II in the present report. Changing the instructions given to the normal *S*s produced more severe retardation in solution than did the presence of brain injury under simple (restricted) instructions in Experiment I.

The protocols revealed certain qualitative differences in any group between the slow *S*s and those who solved the problem quickly. The slow *S*s exhibited behavior akin to the "functional fixedness" described by early investigators of problem solving, notably Goldstein, Wertheimer, Duncker, and Maier. These slow *S*s had great difficulties in overcoming the "obvious" way of using the stimulus objects in favor of an unusual way. However, this qualitative difference between slow and fast *S*s cannot be identified with the difference between brain injured patients and normal controls. The functional fixedness noted was as characteristic of the two slow control *S*s (out of 20) in Experiment I, as it was of the 16 slow brain injured patients (out of 40). In Experiment II, where 20 additional controls worked under more difficult conditions ("less restricted," and "unrestricted" instructions), there appeared a great deal of varied and tentative presolution behavior which resembled in many ways that displayed by the slower brain injured and normal *S*s under simple (restricted) conditions.

---

[3] Even in the normal, the relationship between ability of seeing concealed figures and ability in problem situations of the type here considered seems rather complex (22).

In view of such similarities between slow *S*s with and without brain injury it becomes impossible to interpret our findings as a confirmation of a sharp dichotomy between problem solving behavior before and after cerebral lesions (15, 16, 17, 26, 43). There was nothing in our data, quantitatively or qualitatively, that would preclude the possibility that inefficiency displayed by the brain injured was not of the same kind (though more frequent and persistent) as the inefficiency found among the less competent normal controls.

The presolution behavior of normal as well as brain injured *S*s included striking instances of seemingly irrelevant manipulations, reminiscent of certain abortive attempts of Köhler's chimpanzees in the box stacking problem (35). Thus, several *S*s in each group tried to solve the string problem by opening the scissors as widely as possible, putting one point near the end of one string, and extending the other end in the general direction of the second string. One may compare this act with that of Sultan (Köhler's most gifted chimpanzee) who had reached a point in his presolution period of the box stacking problem where he apparently understood that two boxes were required to reach the suspended lure. Cf. Köhler: "(Sultan took second box) . . . but, instead of placing it on top of the first (box) . . . began to gesticulate with it . . . put it beside the first, then in the air diagonally above and so forth" (35, p. 47). It is important to note that such behavior was not restricted to our brain injured *S*s, but was displayed by some of the normal *S*s in Experiment I, and by many of the normal *S*s in Experiment II. Had this "magical" behavior been restricted to the brain injured group, it might have led to the assumption that problem solving, in the injured, proceeded on a different and "lower" level than in the normal. The facts, however, showed considerable variation of ability for problem solving from individual to individual, irrespective of presence or absence of the cerebral lesion. We would therefore restrict the terms customarily applied to various "levels" of problem solving behavior to a descriptive rather than explanatory use (60).

Levels of problem solving have been assumed to exist by many investigators, notably Goldstein (abstract vs. concrete attitude) (15, 16, 17) and Maier (productive vs. reproductive thinking) (42, 43, 54). As usually stated, these views imply a dichotomy in the manner of solving a problem and predict a descent from higher to lower levels following massive brain injury (15, 43). Our data failed to yield evidence for such a dichotomy between brain injured and normal groups. This might mean that the dichotomy does not exist, or that it cannot be demonstrated by the task or *S*s employed.

It could be argued, for instance, that the groups studied simply did not include a sufficient number of those severely disabled men with brain injuries, who were preferentially tested in previous studies by others (9, 13, 33, 49). Qualitative alterations in the approach to problems may be produced by brain injury, if the extent of the lesion exceeds a critical mass (37, 43), or if the lesion is of an irritative nature (29, 30). Further studies would be necessary to clarify these issues.

The main result of the present study would then be empirical, demonstrating that some difficulties in problem solving on a Maier-type task may appear after cerebral lesions but are not an obligatory result of such lesions. In some, but not all patients, brain injury interfered with problem solving; but the nature of this interference was indistinguishable in its manifestations from the ineptness found in those *S*s in our control group who likewise showed retardation and functional fixedness in their problem solving behavior.

F. SUMMARY AND CONCLUSIONS

Eighty World War II veterans were tested on a modification of the Maier String Problem (41). Twenty veterans had penetrating injuries in frontal regions of the brain, 20 in the parieto-occipital regions (with homonymous field defects); the remaining 40 veterans had peripheral nerve injuries (controls).

Each subject entered a room where two strings were hanging freely from the ceiling. His task was to tie the ends of these strings together, although they were too far apart to permit grasping one while holding the other. The problem could be solved only by attaching an object to one of the strings, thus creating a pendulum, and swinging it toward the other string. In Experiment I (restricted instructions) the only objects available to the subject were provided by the experimenter (in two-minute intervals) in this order: scissors, clothes pin, pulley, yo-yo, sinker. These objects served as a graded series of practical "hints" toward solution. In Experiment II (less restricted and unrestricted instructions) the same objects were available either at two-minute intervals, or simultaneously from the start of the problem, but in either case the subject was permitted to use anything in the room as long as a "pendulum" solution was achieved.

Results show that brain injury in frontal or parieto-occipital areas is less disruptive of problem solving than might have been expected from previous reports. Some men in either of the brain injured groups took longer, and depended on more specific "hints," in achieving solution than did the controls, but group differences were too small to reach significance (Experiment I).

In normal subjects, ability to solve problems of the type employed is more highly correlated with ability to perceive hidden contours (in a modified set of Gottschaldt figures) than with ability to sort to criterion (in a Weigl type sorting test). Neither of these correlations was significant in the brain injured groups.

Patients with injuries involving the anterior lobe substance did not do any worse than those with lesions in the posterior lobe substance. This finding is not compatible with prevailing assumptions that integrity of the frontal lobes is of singular importance for the solution of problems of this nature.

REFERENCES

1. ACH, N. Über die Willenstätigkeit und das Denken. Göttingen: Vandenhoeck & Ruprecht, 1905.
2. BLUM, J. S., CHOW, K. L., & PRIBRAM, K. A behavioral analysis of the organization of the parieto-temporo-preoccipital cortex. *J. Comp. Neurol.*, 1950, **93**, 53-100.
3. BLUM, R. A. The effects of bilateral removal of the prefrontal granular cortex on delayed response performance and emotionality in chimpanzee. *Amer. Psychol.*, 1948, **3**, 237-238.
4. BRICKNER, R. The Intellectual Functions of the Frontal Lobes. New York: Macmillan, 1936.
5. BULBROOK, M. E. An experimental inquiry into the existence and nature of "insight." *Amer. J. Psychol.*, 1932, **44**, 409-453.
6. CHOW, K. L., BLUM, J. S., & BLUM, R. A. Effects of combined destruction of frontal and posterior "associative areas" in monkeys. *J. Neurophysiol.*, 1951, **14**, 59-71.
7. CLAPARÈDE, E. La genèse de l'hypothèse. *Arch. de Psychol.*, 1934, **24**, 1-154.
8. DUNCKER, K. Zur Psychologie des produktiven Denkens. Berlin: Springer, 1935. (Trans. by L. S. Lees as: On problem solving, *Psychol. Monog.*, 1945, **58**, Whole No. 270.)

9. FEUCHTWANGER, E. Die Funktionen des Stirnhirns, ihre Pathologie und Psychologie. *Monog. a. d. Ges. geb. d. Neurol. u. Psychiat.*, (Ed. O. Foerster and K. Williams), Heft 38. Berlin: Springer, 1923.

10. FREEMAN, W. Mass action versus mosaic function of the frontal lobe. *J. Nerv. & Ment. Dis.*, 1949, **110**, 413-318.

11. FREEMAN, W., & WATTS, J. W. Psychosurgery: Intelligence, Emotion and Social Behavior following Prefrontal Lobotomy for Mental Disorders, (2nd ed.). Springfield, Ill.: Thomas, 1951.

12. FULTON, J. F. Functional Localization in the Frontal Lobes and Cerebellum. (The William Withering Memorial Lectures, 1948.) Oxford: Clarendon Press, 1949.

13. GELB, A., & GOLDSTEIN, K. Psychologische Analysen hirnpathologischer Fälle. Leipzig: Barth, 1920 sqq. (Partially translated in, Ellis: *Source Book of Gestalt Psychology*, New York: Harcourt Brace, 1938.)

14. GERMAN, W. J., & FOX, J. C. Observations following unilateral lobectomies. *Res. Pub. Assoc. Nerv. Ment. Dis.*, 1934, **13**, 378-434.

15. GOLDSTEIN, K. After-Effects of Brain Injuries in War. New York: Grune & Stratton, 1942.

16. ———. The mental changes due to frontal lobe damage. *J. of Psychol.*, 1944, **17**, 187-208.

17. GOLDSTEIN, K. & SCHEERER, M. Abstract and concrete behavior: an experimental study with special tests. *Psychol. Monog.*, 1941, **53**, 1-151.

18. GOTTSCHALDT, K. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren: I. *Psychol. Forsch.*, 1926, **8**, 261-317.

19. ———. Über den Einfluss der Erfahrung auf die Wahrnehmung von Figuren: II. *Psychol. Forsch.*, 1929, **12**, 1-87.

20. ———. Der Aufbau des kindlichen Handelns. *Beih. z. angew. Psychol.*, 1933, No. 68.

21. GRANT, A. D., & BERG, E. A. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. Exp. Psychol.*, 1948, **38**, 404-411.

22. GUETZKOW, H. An analysis of the operation of set in problem solving behavior. Ph.D. dissertation, Univ. of Michigan Library, Ann Arbor, Mich., 1947.

23. HALSTEAD, W. C. Preliminary analysis of grouping behavior in patients with cerebral injury by the method of equivalent and non-equivalent stimuli. *Amer. J. Psychiat.*, 1940, **96**, 1263-1294.

24. ———. Brain and Intelligence: A Quantitative Study of the Frontal Lobes Chicago: Univ. Chicago Press, 1947.

25. HANFMANN, E., & KASANIN, J. A method for the study of concept formation. *J. of Psychol.*, 1936, **3**, 521-540.

26. HARLOW, H. F. The effect of large cortical lesions on learned behavior in monkeys. *Science*, 1950, **112**, 428.

27. HARLOW, H. F., & SETTLAGE, P. H. Effect of extirpation of frontal areas upon learning performance of monkeys. *Res. Publ. Assoc. Nerv. & Ment. Dis.*, 1948, **27**, 446-459.

28. HEBB, D. O. Intelligence in man after large removals of cerebral tissue: report of four left frontal lobe cases. *J. Gen. Psychol.*, 1939, **21**, 73-87.

29. ———. Man's frontal lobes: a critical review. *Arch. Neurol. & Psychiat.*, 1945, **54**, 10-24.

30. HEBB, D. O., & PENFIELD, W. Human behavior after extensive bilateral removal from the frontal lobes. *Arch. Neurol. & Psychiat.*, 1940, **44**, 421-438.

31. HOBHOUSE, L. T. Mind in Evolution. (Rev. Ed.) London: Macmillan, 1915.

32. HUMPHREY, G. Thinking: An Introduction to its Experimental Psychology. New York: Wiley, 1951.

33. KLEIST, K. Gehirnpathologie. Leipzig: Barth, 1934.

34. KLÜVER, H. Behavior Mechanisms in Monkeys. Chicago: Univ. Chicago Press, 1933.

35. KÖHLER, W. The Mentality of Apes. New York: Harcourt Brace, 1925.

36. LANDIS, C., ZUBIN, J., & METTLER, F. A. The functions of the human frontal lobe. *J. of Psychol.,* 1950, **30**, 123-138.

37. LASHLEY, K. S. Brain Mechanisms and Intelligence. Chicago: Univ. Chicago Press, 1929.

38. LEWIS, D., & BURKE, C. J. Further discussion of the use and misuse of the chi-square test. *Psychol. Bull.,* 1950, **47**, 347-355.

39. LINDLEY, E. H. A study of puzzles with special reference to the psychology of mental adaptation. *Amer. J. Psychol.,* 1897, **8**, 431-493.

40. MAIER, N. R. F. Reasoning in humans: I. On direction. *J. Comp. Psychol.,* 1930, **10**, 115-143.

41. ———. Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *J. Comp. Psychol.,* 1931, **12**, 181-194.

42. ———. Reasoning in humans: III. The mechanism of equivalent stimuli and of reasoning. *J. Exp. Psychol.,* 1945, **35**, 349-360.

43. ———. A further analysis of reasoning in rats: II. The integration of four separate experiences in problem solving. III. The influence of cortical injuries on the process of "direction." *Comp. Psychol. Monog.,* 1938, **15**, 1-85.

44. MAIER, N. R. F., & REININGER, H. W. A Psychological Approach to Literary Criticism. New York: Appleton, 1933.

45. MALMO, R. B. Interference factors in delayed response in monkeys after removal of frontal lobes. *J. Neurophysiol.,* 1942, **5**, 295-308.

46. ———. Psychological aspects of frontal gyrectomy and frontal lobotomy in mental patients. *Res. Pub. Assoc. Nerv. Ment. Dis.,* 1948, **27**, 537-546.

47. METTLER, F. A. (Ed.) Columbia-Greystone Associates: Problems of the Human Brain: I. Selective Partial Ablation of the Frontal Cortex. New York: Hoeber, 1949.

48. PENFIELD, W. & EVANS, J. The frontal lobe in man: a clinical study of maximum removals. *Brain,* 1935, **58**, 115-133.

49. POPPELREUTER, W. Die psychischen Schädigungen durch Kopfschuss im Kriege 1914-16. Leipzig: Voss, 1917.

50. RUGER, H. A. The psychology of efficiency: an experimental study of the processes involved in the solution of mechanical puzzles and in the acquisition of skill in their manipulation. *Arch. Psychol.,* 1910, **2**, Whole No. 15.

51. RYLANDER, G. Personality Changes After Operations on the Frontal Lobes. London: Oxford, 1939.

52. ———. Mental changes after excision of cerebral tissue: A clinical study of 16 cases of resections in the parietal, temporal, and occipital lobes. *Acta Psychiat. et Neurol.,* 1943 (Suppl.), **20**, 1-81.

53. ———. Personality analysis before and after frontal lobotomy. *Res. Pub. Assoc. Nerv. & Ment. Dis.,* 1948, **27**, 691-705.

54. SELZ, O. Zur Psychologie des produktiven Denkens und des Irrtums. Bonn: Cohen, 1922.

55. SETTLAGE, P., ZABLE, M., & HARLOW, H. F. Problem solutions by monkeys following bilateral removal of the prefrontal areas: VI. Performance on tests requiring contradictory reactions to similar and to identical stimuli. *J. Exp. Psychol.,* 1948, **38**, 50-65.

56. TEUBER, H. L. Some observations on the organization of higher functions after penetrating brain injury in man. In: *The Biology of Mental Health and Disease.* New York: Hoeber-Harper, 1952.

57. TEUBER, H. L., BATTERSBY, W. S. & BENDER, M. B. Performance of complex visual tasks after cerebral lesions. *J. Nerv. & Ment. Dis.,* 1951, **114**, 413-429.

58. ———. Effects of cerebral lesions on intellectual-functioning in man. *Fed. Proceedings,* 1952, **11**, 161.

59. THURSTONE, L. L. A Factorial Study of Perception. Chicago: Univ. Chicago Press, 1944.

60. WEAVER, H. E., & MADDEN, E. H. "Direction" in problem solving. *J. of Psychol.,* 1949, **27**, 331-345.

61. WEIGL, E. On the psychology of so-called processes of abstraction. *J. Abn. & Soc. Psychol.,* 1941, **36**, 3-33.

62. WERTHEIMER, M. Über Schlussprozesse im produktiven Denken (1920). In: *Drei Abhandlungen zur Gestalttheorie.* Berlin: Erlangen, 1925. (Pp. 164-189.)

63. ———. Productive Thinking. New York: Harper, 1945.

64. WILCOXON, F. Some Rapid Approximate Statistical Procedures. Stamford: American Cyanamid, 1949.

Department of Neurology
New York University College of Medicine
477 First Avenue
New York City 16

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 353-360.

THE PERFORMANCE OF ALBINO MICE IN THE MAZE SITUATION WITH THE UTILIZATION OF LIGHT AS MOTIVATION AND ITS RELATIVE ABSENCE AS REINFORCEMENT*

Department of Psychology, The George Washington University

WILLARD E. CALDWELL AND H. LYNN WOMACK[1]

A. INTRODUCTION

The problem of utilizing different kinds of motivation and reward is a very important one in the field of comparative psychology and animal learning. John P. Flynn and E. A. Jerome (3, 4), have performed some interesting experiments utilizing light avoidance as motivation in the albino rat. Caldwell and Mosman (1) utilized temperature as a motivation and Caldwell and Sandler (2) utilized gaseous formaldehyde as a motivation. The subjects in both these experiments were albino mice.

B. STATEMENT OF THE PROBLEM

The purpose of this experiment is to further investigate the motivation utilized by Flynn and Jerome by varying the animal and utilizing mice[2] who would be subjected to the more conventional maze situation consisting of six brightly lighted blind alleys and a goal box which would be almost totally dark. Some of the objectives may be more specifically stated as: What is the nature of the learning curve for a group of albino mice subjected to a brightly lighted maze with a dark goal box as motivation and reinforcement, as compared with the learning curve for a control group of mice subjected to the same maze pattern but in a maze painted black throughout and without being subjected to the difference in light or to any other known form of motivation or reinforcement?

C. APPARATUS

The apparatus for this experiment consisted of a wire maze which had six blind alleys. The entire maze, with the exception of the entrance

---

*Received in the Editorial Office on December 18, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The authors wish to express their appreciation to Messrs. H. N. Alexander, D. R. Gruver, F. W. Lane, J. E. Ridgeway, and E. P. Sheehy for assistance received in conducting the experiment and in the preparation of this paper.

[2] At the present time plans are being drawn to experiment on the rôle of light avoidance in the gold fish and the salamander.

and goal boxes, but including the box to which the maze was attached, was painted with aluminum paint. A light bulb of 300 watts, above which was a white enamel reflector, was suspended 24 inches above the entire maze. The entrance compartment and the goal box were made of clear plastic. The goal box was covered with ordinary black oil cloth and had a sliding door which could be closed by the experimenter. A flap, made of black oil cloth, could be pulled down over the closed door and the goal box thus made almost, if not completely, dark. The plastic entrance compartment had two sliding doors, the first opened in placing the mice in the compartment, and the second opened when the experimenter was ready for the animal to run the maze. One thermometer was kept at the entrance to the maze and another in the goal box for both the experimental and control groups. The maze, which rested on a small table, was surrounded by black curtains which were suspended from the ceiling of the laboratory. The area enclosed by the black curtains was lighted by a single fluorescent unit containing four tubes which were 36 inches long and each had a strength of 20 watts. The above maze was used on the experimental group. For the control group the same maze was used but the entire maze and the box to which the maze was attached was painted black. The transparent entrance compartment used with the experimental group was used with the control group and a black goal box was used, the goal box having a cover of black screen wire. See Figure 1 for a picture of the apparatus.

D. PROCEDURE

The mice used in this experiment were four weeks old Swiss white mice from the Dovel Laboratory Strain. In both the experimental and control groups 12 mice were used.[3] When as much as 2/3 of an animal's body went into a blind alley, that was counted as an error. When the animal would turn back in the maze, this was counted as a retrace until he returned to the point from which the retrace began. When the animal entered a blind alley during a retrace, this was counted as an error as well as a retrace. Time was recorded with a stop watch from the time the door of the entrance compartment leading into the maze was opened until the door to the goal box was closed. The animals were left in the goal box for five minutes. Temperatures at the entrance compartment and the goal box were recorded at the beginning of each trial and when the animals were

---

[3] One animal died at the conclusion of Trial 8 and another at the conclusion of Trial 15 in the control group.

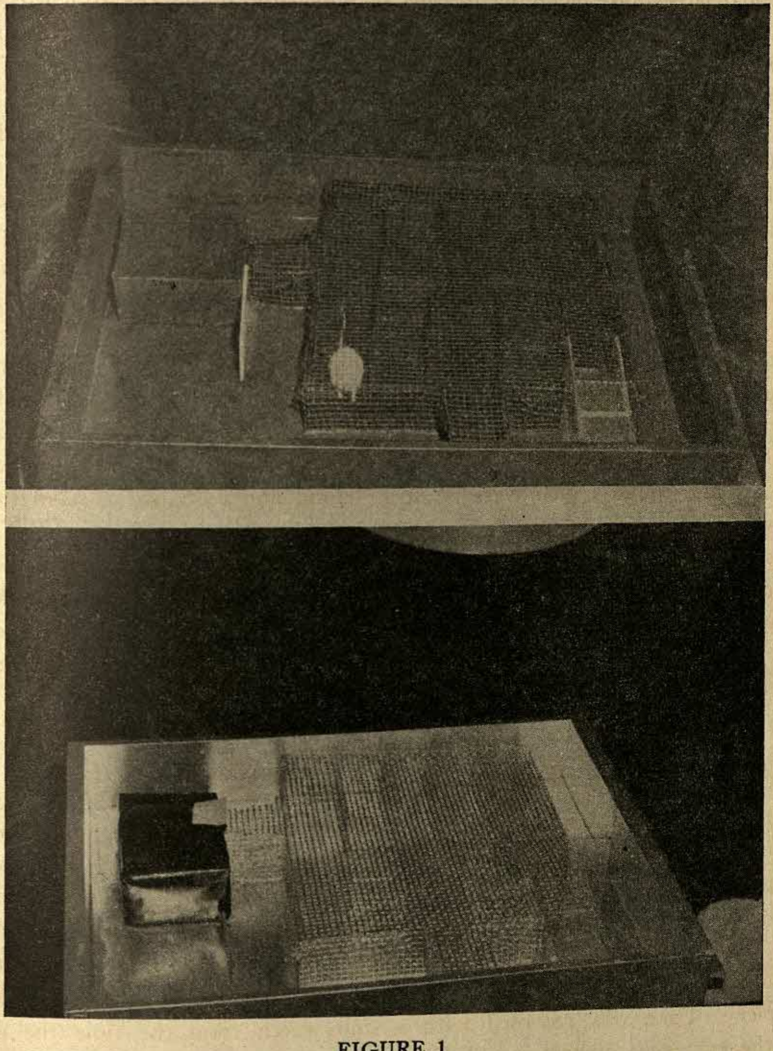


FIGURE 1

The upper maze, utilized for the control group with no difference in light between goal box and maze, was painted black. The lower maze, utilized for the experimental group, was painted with aluminum paint.

removed from the goal box. Two experimenters ran the animals in the experimental and control groups. During the first part of the experiment both observers recorded errors for each run and these were checked to make certain that both observers were using identical criteria for recording errors and retraces. The animals were run every day for 20 days during the hours of 6:00 P.M. and 12:00 A.M. Each animal was kept separately in an individual glass jar, in which food and water were available at all times. At the conclusion of each daily trial, the individual jars were cleaned and refilled with fresh sawdust and food. The temperature for the entire laboratory was checked throughout the experiment and was found to vary between 74 and 81 degrees Fahrenheit. There were no sharp fluctuations as the heat in the laboratory was thermostatically controlled.

E. RESULTS

Time data (Table 1) reveal significant *t* values between control and experimental groups at the 5 per cent level, in favor of the experimental group, on Trials 8, 9, and 10; and in favor of the control group on Trial 20.

Error data (Table 2) show significant *t* values at the 5 per cent level for Trials 4, 8, and 9 in favor of the experimental group; and Trials 14, 16, and 18 in favor of the control group.

Learning curves, representing mean time in seconds, for the experimental and control groups may be seen in Figure 2. Curves for errors may be seen in Figure 3.

F. DISCUSSION OF RESULTS

If an examination is made of Table 1 and Figure 2, it will be noted that the experimental group is, without exception, taking a longer time to run the maze up to Trial 12. At Trial 12 the mean time per trial for the experimental group begins to decrease and for the remaining 8 trials the experimental group requires less time per trial than does the control group. It is interesting to note that the peak of the curve for the experimental group (Figure 2) is between Trials 8, 9, and 10; and that the difference for these trials are statistically significant.

If an examination is made of Table 2 and Figure 3, it will be noted that the experimental group is, without exception, making more errors than is the control group up to Trial 12; but after Trial 12 the control group makes more errors per trial than does the experimental group. It is also interesting that the peak of the error curve for the experimental group is around Trials 8 and 9 and that the difference for both these trials are statistically significant.

TABLE 1

MEAN TIME IN SECONDS FOR EXPERIMENTAL AND CONTROL GROUPS*

	Trials									
	1	2	3	4	5	6	7	8	9	10
	Mean Time in Seconds									
$\overline{X}e$	510.66	448.33	573.50	607.83	673.58	582.83	569.16	1,059.66	790.66	1,181.75
$\overline{X}c$	130.66	180.00	76.75	51.33	71.91	192.08	208.00	81.91	124.27	112.72
σe	939.35	718.44	947.77	918.66	945.33	958.15	960.60	1,272.20	772.25	1,248.87
σc	103.37	189.23	125.22	60.86	72.25	219.32	193.27	67.35	143.86	120.53
D	380.00	268.33	496.75	556.50	601.67	390.75	361.16	977.75	666.39	1,069.03
t	1.33	1.22	1.72	2.00	2.10	1.31	1.22	2.54	2.80	2.82

	Trials (*continued*)									
	11	12	13	14	15	16	17	18	19	20
	Mean Time in Seconds									
$\overline{X}e$	634.08	431.50	409.33	408.66	369.25	437.33	597.75	534.25	502.66	184.50
$\overline{X}c$	332.09	441.81	818.81	888.00	1,035.36	921.80	907.60	1,381.70	1,201.30	1,048.80
σe	561.98	426.09	367.00	350.42	330.03	601.41	670.96	759.55	813.07	180.24
σc	431.10	509.52	1,049.10	1,286.93	1,262.91	1,144.45	1,010.53	1,250.31	1,308.07	1,078.97
D	301.99	10.31	409.48	479.34	666.11	484.47	309.85	847.45	698.64	864.30
t	1.38	.05	1.17	1.14	1.61	1.14	.78	1.78	1.39	2.37

*For Trials 1 through 8, at the 5 per cent level of significance, a t value of 2.074 is required—degrees of freedom being 22. For Trials 9 through 15, at the 5 per cent level of significance, a t value of 2.080 is required—degrees of freedom being 21. For Trials 16 through 20, at the 5 per cent level of significance, a t value of 2.086 is required—degrees of freedom being 20.

TABLE 2

MEAN ERRORS RECORDED FOR EXPERIMENTAL AND CONTROL GROUPS*

	Trials									
	1	2	3	4	5	6	7	8	9	10
					Mean Errors					
$\overline{X}e$	20.91	12.08	17.41	25.16	45.08	33.16	27.91	61.16	44.83	36.75
$\overline{X}c$	7.58	14.08	6.33	5.08	8.58	13.25	20.00	10.41	11.36	9.90
σe	29.42	10.53	12.25	17.08	74.27	36.70	28.95	80.55	46.17	33.99
σc	6.61	6.88	6.47	3.94	5.85	9.75	12.45	8.60	13.64	10.06
D	13.33	2.00	11.08	20.08	36.50	19.91	7.91	50.75	33.47	26.85
t	1.45	.52	2.65	3.81	1.62	1.74	.83	2.07	2.29	2.50

	Trials (*continued*)									
	11	12	13	14	15	16	17	18	19	20
					Mean Errors					
$\overline{X}e$	23.50	15.58	15.33	9.58	8.08	8.41	25.58	12.25	14.00	10.16
$\overline{X}c$	13.45	19.63	29.72	26.72	25.54	23.90	32.00	33.50	25.40	19.80
σe	25.25	15.72	12.39	12.25	12.03	7.44	24.85	12.66	15.40	9.53
σc	15.78	19.66	32.62	18.75	25.05	16.40	14.32	23.00	22.20	14.68
D	10.05	4.05	14.39	17.14	17.46	15.49	6.42	21.25	11.40	9.64
t	1.10	.51	1.31	2.45	2.00	2.62	.72	2.48	1.30	1.70

*For Trials 1 through 8, at the 5 per cent level of significance, a t value of 2.074 is required—degrees of freedom being 22. For Trials 9 through 15, at the 5 per cent level of significance, a t value of 2.080 is required—degrees of freedom being 21. For Trials 16 through 20, at the 5 per cent level of significance, a t value of 2.086 is required—degrees of freedom being 20.

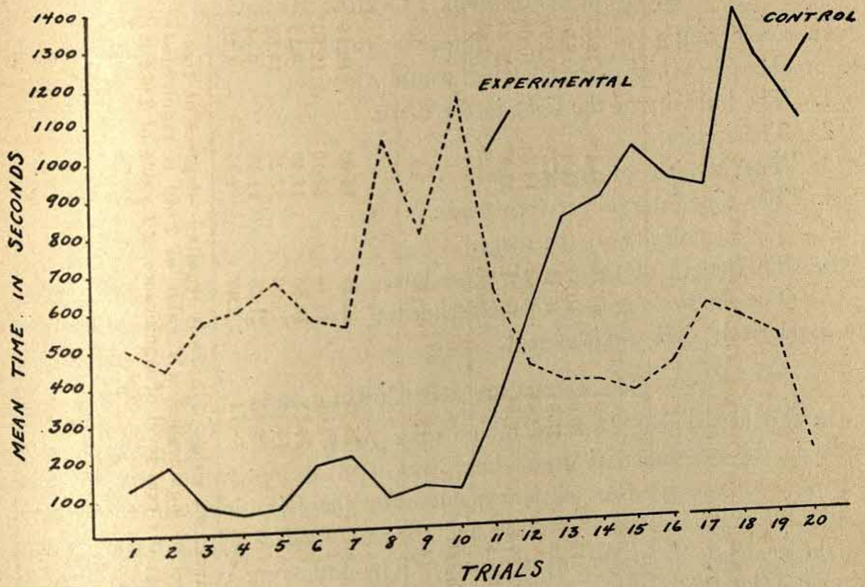


FIGURE 2
LEARNING CURVES REPRESENTING MEAN TIME IN SECONDS FOR BOTH EXPERIMENTAL AND CONTROL GROUPS

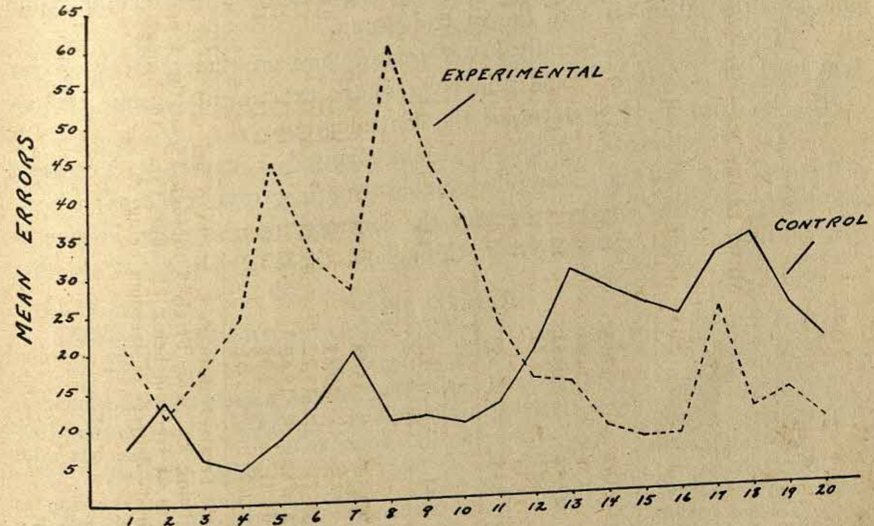


FIGURE 3
LEARNING CURVES REPRESENTING MEAN ERRORS FOR BOTH EXPERIMENTAL AND CONTROL GROUPS

G. SUGGESTIONS FOR FURTHER RESEARCH

In considering possibilities for future research, attention might be given to problems in which the following would vary:

1. The intensity of the light in the maze.
2. The species.
3. The number of trials.
4. The time interval between trials.
5. The complexity of the maze.
6. The length of time in the goal box.
7. The environment of the experimental group, such as running them in a relatively dark environment.

H. SUMMARY AND CONCLUSIONS

In this experiment 24 sated albino mice were utilized. For the 12 animals in the experimental group light was used as motivation and its relative absence was used as reinforcement. For the 12 animals in the control group the maze was painted black and the same degree of light was present in the goal box as in the maze proper. The following conclusions may be drawn from this experiment:

1. There was a statistically better performance, at the 5 per cent level, for time scores on Trials 8, 9, and 10 in favor of the experimental group and on Trial 20 in favor of the control group.

2. There was a statistically better performance on errors, at the 5 per cent level, on Trials 4, 8, and 9 in favor of the experimental group and on Trials 14, 16, and 18 in favor of the control group.

3. The experimental group in general showed a curve of increasing errors and time and then showed a gradual drop in errors and time; whereas the general shape of the curve for the control group seems to be one of gradual increase of time and errors throughout the 20 trials.

REFERENCES

1. CALDWELL, W. E., & MOSMAN, K. F. The rôle of temperature change as reinforcement. *J. of Psychol.*, 1951, **32**, 231-239.
2. CALDWELL, W. E., & SANDLER, H. M. The rôle of gaseous formaldehyde as reinforcement in maze learning in albino mice. *J. of Psychol.*, 1952, **33**, 47-56.
3. FLYNN, J., & JEROME, E. A. Learning in an automatic multiple-choice box with light as incentive. *J. Comp. & Physiol. Psychol.*, 1952, **45**, 336-340.
4. JEROME, E. A., & FLYNN, J. P. A multiple choice box using light aversion as motivation. Bethesda, Md.: Naval Medical Research Institute, 1950. Project NM 000 019.01.01.

Department of Psychology
The George Washington University
Washington 6, D. C.

Published as a separate and in *The Journal of Psychology*, 1953, **35**, 361-370.

GROUP MEMORY OF MEANINGFUL MATERIAL*[1]

Center for International Studies, Massachusetts Institute of Technology

HOWARD V. PERLMUTTER[2]

A. PROBLEM

There has been no reported research in small-group psychology that provides some understanding of the properties of group "memories." The difficulty involved in measuring group phenomena and group memory in particular has seemed formidable. In his classic study, *Remembering* (1) Bartlett wrote:

> If we wish to know whether a man remembers anything, we ask him; and we generally take his word for it, even when he can give us but little detail. We cannot adopt this easy way with the social group because, if it has means of self-expression other than its contemporary movements and its social organization, we do not know that means. . . .

While Halbwachs (2, 3) has suggested that families and institutions retain vivid impressions of events and that no particular member of the group can recall adequately the total shared experience of that group without assistance from the other group members, he, too, provided no experimental evidence. His keen observations should appeal to the common experience of group psychologists who, in their examination of the problems of group growth, recognize that it is necessary to speak of a past shared experience of a group that explains in part the present behavior of that group. But psychologists generally feel a reluctance to attribute individual functions to groups. It was suggested that some experiments primarily associated with individuals can be repeated with groups. Horowitz and this writer (4) have considered the methodological basis for this research elsewhere. Rather than continue further controversies about "group mentalities," empirical research is considered appropriate at this stage of development in small group psychology.

This study had three general areas of interest:

1. The products of a group discussion wherein the members of a group are asked to produce a collective "memory." The technique for evoking the group memory, called the *Method of Agreement*, was reported previously in a study of group learning (6).

---

*Received in the Editorial Office on December 20, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Selections from the dissertation "A Study of Group and Individual Memory Products" submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, Kansas University, September, 1952.

[2] Also at Department of Psychology, Boston University.

2. A comparison of the principles of group memory change with those of individual memory change. One of Bartlett's experiments using his *Method of Repeated Reproduction* was reported using the same story. He had concluded:

> *a*. Accuracy of reproduction of the original story, in a literal sense, was rare.
>
> *b*. The general form of the story was very persistent once the first recall occurred.
>
> *c*. Omission of detail, simplification of events, structure, and transformation of items into more familiar detail continued through subsequent recalls.

So we desired to find out if these conclusions apply as well to group-memory products.

3. The data relevant to the question: to what extent is the group memory-product derivable from the memories of the individual members?

B. PROCEDURE

Sixteen experimental groups, comprised of eight two-person groups and eight three-person groups, were used. Four of the three-person groups were composed of females; four of males. The three-person groups were divided in the same manner. A control group of 10 individuals who worked in isolation was also used.

The subjects were college students at the University of Kansas. No attempt was made to equate the experimental and control subjects in terms of intelligence or other personality factors. Most subjects knew the other members of the group "by sight only."

The groups and individuals met on two consecutive days. On the first day after a warming-up period, the story "War of the Ghosts" was read twice. The group and individuals were asked to recall the story 15 minutes after hearing it and again 24 hours later. On the second day, the members of the group first recalled the story separately by writing their conceptions of the story on a sheet of paper. All group recalls were dictated to a recorder (not a member of the group) who wrote the group's product on a bulletin board in full view of all members of the group.

Two observers noted the time required to recall the story, the degree of agreement the group expressed on each portion of the story it recalled, and then rated the group process. The group and individual memory products were scored in the following fashion. The original story was divided into 52 meaningful segments. The scorer rated each segment on a 10-point scale for the degree to which the group or isolated individuals duplicated in meaning and style the original segment. The total score was the sum of the ratings for all segments. The maximum possible score was 520. Inter-scorer agreements on units was very high.

C. Discussion and Results

1. *The Method of Agreement*

The method of agreement required a collective agreement on each portion of the story remembered by the group. Observers ranked each decision according to the degree of agreement among the members of the group concerning an adapted portion of the story. The average score by the observers for all experimental groups corresponded to the scale rating: "Group in general agreement; one person overtly, the others tacitly agree or do not disagree." This coupled with the results of another study (6) seems to indicate the method of agreement induced the conditions it purported to induce.

2. *Comparison of Group and Individual Memory Products*

a. Total score and time required for recall. Quantitative comparisons of group and individual memory were made with respect to the amount of material recalled and the time required to recall. Table 1 summarizes the results.

TABLE 1
Comparison of Group and Individual Memory Products

		Session I				*Session II*			
	N	Ave. total score	*SD**	Ave. time for recall (minutes)	*SD**	Ave. total score	*SD**	Ave. time for recall (minutes)	*SD**
3-person Groups	8	311	46.8	29.7	8.1	303	39.5	18.7	3.8
2-person Groups	8	299	46.7	25.5	9.3	288	51.1	17.4	3.7
Individuals	10	263	69.5	14.1	6.5	254	66.9	12.6	4.1

In these comparisons, the following differences were found to be significant. (*a*) With respect to total scores (maximum score is 520) no differences are statistically significant although the results favor the groups. The standard deviations of the individuals' scores are nearly twice those of the three person groups, indicating that the pressures toward conformity may be operating so that group products are less variable than individual products. (*b*) Individuals require less time than both two and three person groups at at least .05 level of significance in both sessions.

b. Applicability of principles of memory change. With respect to the applicability of the principle of memory change, found by Bartlett, group and individual products were compared. Bartlett had found first of all that *accuracy of reproduction, in a literal sense, will be rare.* The evidence supports these statements. Very few of the groups produced faithfully more than six or seven segments of the original story. The last sentence of the story, "He was dead," was the one most often recalled. In the first session, all groups recalled this sentence, and seven out of 10 individuals (who worked alone), recalled this exact phrasing. The other three wrote a close version, "He died." The other 12 segments were recalled by smaller percentages of the groups and individuals. The second half of the first sentence was very frequently recalled exactly ("went down to the river to hunt seals"). The dramatic phrase, "something black came out of his mouth," was also very frequently recalled in its exact form. Since the number of exact segments produced were few, obviously distortions of the original story were the rule. Particular segments, however, were distorted more frequently than others. Segments with peculiar phraseology as "Now canoes came up," "And they said, 'What do you think?'," "Behold, I accompanied the ghosts," and "And he told everybody and said," are commonly distorted by both individuals and groups.

The second conclusion taken from Bartlett stated: *the general form of the story will be very persistent, once the first recall occurs.* The evidence supports this statement. Generally, the characteristic style of the first recall resembles greatly that of the second version. The quantitative results show that the scoring of a majority of the segments remains the same[2] for both sessions. On the average, about 63 per cent of the segments for groups and about 76 per cent for individuals receive exactly the same score for the recalls of both sessions. Changes in presentation do occur as, for example, "river" becomes "seashore" in the second session. The construction of the second sentence is changed. "When they heard shouts they thought—" becomes, "thinking it was a war party. . . ." But distortions such as "shouts for noises" in the original story, "afraid of being killed" for "I might be killed," persist from Session I to Session II. In another example, the group wrote in both sessions "One canoe containing five men," while the original reads, "There were five men in the canoe."

Bartlett's third conclusion states: *Omission of detail, simplification of events and structure, and transformation of items into more familiar detail will continue during the two sessions.* The average total recall score for

---

[2] To receive an equal score on both days, a segment had to remain almost exactly the same in *both* style and content.

individuals and groups was lower on Session II. This decrement in performance comes from more omission of details from the original story. Redundant phrases like "He told it all" and "Now canoes came up" were frequently omitted since they were not required for a coherent story. Several phrases were either completely remembered or totally omitted; very few partial scores were noted for those phrases. Examples of such segments are "His face became contorted" and "Many were killed."

Summarizing, there are on the average quantitative differences with respect to applicability of the principles of memory change. Some particular groups are however identical along all dimensions to individuals who work alone. We pose as a hypothesis worthy of future investigation: there are general principles that describe both group-qua-group and individual-qua-individual behavior. Groups, depending upon their size and individuals are the quantities that fit in the general statements.[3]

3. *Derivability of Group Memory Products From Member Memory Products*

It has been hypothesized that group memory products are not derivable from the individual recalls of the members as determined in isolation. One derivation could be: some correct content will be found in the group memory product that is not in any member's recall. These data were taken from the dictated recalls of the groups and the written recalls of the *members* of these groups that preceded the second group recall. Each memory product was examined to find those instances in which a particular segment in the group memory product of Session II was not found in the memory product of any individual member. There is hardly any evidence for the extreme position, the *content* of group memory products is unique and not related to the content of individual member recalls.

Still another hypothesis could be: some correct content found in all or some of the individual recalls per se will not be found in the group memory product. The evidence relating to this hypothesis is strong. Since the group had to discuss and decide on a collective product, each individual's contribution was usually altered in some way. In all groups some individuals had a higher point score on various single segments than the group of which they were a member; and in some groups there were individuals who received a higher total score than the group. This means that they very frequently received higher scores on many segments.

---

[3] The sociologist Parsons (5) has suggested this in recent publications, where he uses the term "homology" to refer to certain formal identities between individuals and groups.

In the majority of cases, the group recall score is better than or equal to the best score of an individual member on each of the 52 different segments, but a significant percentage of instances finds superiority of the individual member to the group. Table 2 gives the results.

TABLE 2
COMPARISON OF BEST INDIVIDUAL SCORE WITH GROUP SCORE
(In average per cent of segments)

	B.I. more than G	B.I. equals G	B.I. less than G
3-person groups	12.2%	48.3%	39.5%
2-person groups	17.3%	50.8%	31.9%

These tables, however, give only those cases where *correct* content in an individual recall is or is not included in the group product. Frequently, although it is difficult to ascertain pure cases, incorrect material in an individual product influences greatly the form of the group product, in some cases to the extent that correct material is excluded.

More evidence was found by making a crude analysis to determine the kinds of relationship the individual member's product *per se* had to the group product. Three categories were chosen: (*a*) cases where sentences *in their exact wording* could be found in at least one individual member's product, termed "logically derivable"; (*b*) cases where a *mixture* of two or more members' materials could logically have produced the group product, termed "logical synthesis"; (*c*) cases where, strictly speaking, one could not derive from any logical summation the group product in its final form, "not logically derivable."

Table 3 gives the results of a sampling of 200 statements from two-person group recalls and 200 from three-person group recalls. These were compared with the scores of their respective members' individual recalls and categorized according to the above system. The results are given in percentages. The "not derivable" category, however, includes many cases where differences in style of presentation of the individual members made derivation of the group products very difficult. For example, some words were left out because individual members preferred a direct quote to an indirect quote. The next

TABLE 3
RELATIONSHIP OF GROUP PRODUCT TO INDIVIDUAL PRODUCTS
(According to three given categories)

	Logically derivable	Logical synthesis	Not logically derivable
3-person groups	36.2%	30.6%	33.2%
2-person groups	27.4%	17.1%	55.5%

section considers this case more fully. About one-third of the sentences in three-person group products (Session II) and over one-half in two-person group products were not, strictly speaking, derivable from the individual members' recalls.

One final "derivation" states: the "style" of the group product will be different from the "style" of the individual recalls per se. The style or phraseology of the group product is generally different from a complicated synthesis of the style of the individual member products. One implication of the "not derivable" category was that while the group style persisted from the group product on Day I to Day II, it was not directly related to the individual member style. Several more random examples were chosen, illustrating various degrees of congruence between individual member "style" and group "style" of presentation. This will serve as well to define operationally our use of the term "style." It is to be noted that while the content of the group product and individual product may be the same, the manner of expressing this content may differ. (Table 4). The group used the past tense.

TABLE 4

Group	*Sess. II* Group product	Member products
XI	The young man knew he had been shot but he did not feel sick.	A. He did not feel sick but they said he had been shot. B. One Indian said, "I am shot but I feel no pain."

Individual *A* used the past tense. But Individual *B*, who recalled it more exactly, used a direct quote. Another example illustrates how the "style" of one member was utilized, while the other member influenced the group product to a negligible degree. (Table 5).

TABLE 5

Group	*Sess. II* Group product	Member products
XVI	"My relatives won't know where I am."	A. His relatives wouldn't know where he was. B. "My folks would become worried because they wouldn't know where I went."

Individual *C* approximated the group style most directly since Individuals *A* and *B* used an indirect quote (Table 6). The group omitted the correct phrase "they said" even though all three people included it.

TABLE 6

Group	*Sess. II* Group product	Member products
III	"There are arrows in the boat."	*A.* The warriors said there were arrows around. *B.* The warriors told him they would supply the arrows. *C.* They said, "There are arrows in the boat."

The evidence relating to the complicated relation of individual member products to group products which was given previously, while difficult to quantify, seems cogent. There is a possibility that complex formulae could be constructed to account for the variety of factors operative in synthesizing the individual products. Generally, it is likely that the most correct portions of the individual products will be included. This is due to the fact that when one member suggests a correct segment, other members are likely to "recognize" it even though they were unable to recall that particular portion of the story when they were working alone. But since in eight of the 16 groups there was at least one member who remembered *more* than the group, the limitations of this generalization can easily be seen. In those eight groups one would not be able to predict with any degree of certainty when a correct portion of a member's story would be omitted.

On a larger view, the results suggest that while it is interesting to attempt logical or psychological derivation of group products from individual member products and interpersonal relations, group products can be treated in their own right. Indeed, *some* principles of product change can be formulated without measurement of individual member memory.

D. Some Next Steps

The purpose of this study has really been to open up possibilities for future research and theorizing in group-qua-group psychology. The instructions can be used for research in group learning, memory, perception, motivation, attitude, and personality studies.

Within the area of "group memory," many specific experiments are suggested by this one. What, for example, are the experiences groups tend to remember or forget? What is the relationship between such independent variables as authoritarian and democratic atmosphere, coöperation-competition, cohesiveness, frustration, upon the recall of a variety of kinds of material, controversial or non-controversial? The repetition of most experiments in individual learning is probably possible; the development of a group-qua-group behavior theory is another realistic possibility.

E. SUMMARY

Sixteen experimental groups, eight of three persons and eight of two persons, four control groups of three persons, and 10 individuals who worked in isolation, were used as subjects in an experimental investigation of memory, repeating with both groups and individuals the procedure used by Bartlett.

The results show some superiority, though not statistical, of groups to individuals in amount of recall, in number of contributions required to produce a product, and in the time required for recall. Members in three-person groups tend to recall better alone than either individuals in isolation or members of two-person groups.

The evidence supports the three generalizations of memory change: (*a*) accuracy of reproduction of the original story, in a literal sense, will be rare; (*b*) the general form of the story will be very persistent, once the first recall occurs; (*c*) omission of detail, simplification of events, structure and transformation of items into more familiar detail will continue during the subsequent recalls.

The relative independence of the group product from the individual memory has received some, but not overwhelming, support. In many specific cases it was possible to see unique individual influences in the group product. Many portions of the individual memory-product determined in isolation appeared in practically untouched form in the group product. But in every group product, there were numerous other portions of the recall that were not logically derivable from individual member written recalls. The somewhat intangible factor of style of group product is certainly not obviously related to the styles of the members' recalls.

It was concluded that as long as some aspects of the group product maintain their uniqueness and as long as the group product is rather easily collected, these kinds of data seem to be of value.

Suggestions for further research were made.

REFERENCES

1. BARTLETT, F. Remembering. London: Cambridge University Press, 1932.
2. HALBWACHS, M. Les Cadres Sociaux de la Memoire. Paris: Alcan, 1935.
3. ———. La Memoire Collective. Paris: Presses Universitaire de France, 1950.
4. HOROWITZ, M. W., & PERLMUTTER, H. V. The concept of the social group. *J. Social Psych.*, (in press).

5. PARSONS, T., & SHILS, E. A., *Eds.* Towards a General Theory of Action. Cambridge: Harvard Univ. Press, 1951.

6. PERLMUTTER, H. V., & MONTMOLLIN, G. Group learning of nonsense syllables. *J. Abn. & Soc. Psychol.*, 1952, **47**, 762-769.

Center for International Studies
Massachusetts Institute of Technology
50 Memorial Drive
Cambridge 39, Massachusetts

Published as a separate and in *The Journal of Psychology,* 1953, **35**, 371-378.

PHILOSOPHIC ASSUMPTIONS IN OPERATIONAL PSYCHOLOGY*

Department of Psychology, University of Kentucky

EDWARD NEWBURY

Despite considerable pretensions among operationists in psychology of avoiding philosophic complications (8, 13, 15), their disagreements with introspectionists and among themselves, sometimes in open metaphysical terms (1, 2, 5, 15), suggest the importance of investigating the significance of some of their explicit or implicit epistemological and ontological assumptions.

A. THE PUBLIC NATURE OF SCIENCE

As a means of escaping philosophic distinctions between objective and introspective psychology, much reliance is placed by Stevens (11) and Boring (2) on the public nature of science. About this there can be no argument in science as a social institution. But Boring argues that private experience "can be, and in science must be, published, and the operation of publication is physiological and therefore physical, so that the means of publication come to provide a physical operational definition of experience. In this way experience inevitably loses its privacy, its immediacy, and its nonphysical character" (2, p. 460).

The crux of this argument seems to be what in science must be public. Pratt recognizes this (9, p. 62). He distinguishes between experience, which is private, and the operation of report, which is public (pp. 99f, 102, 106). The public report is not the field of observation but a communication.

From the operational standpoint one can see that the operation of observation is distinguishable from the operation of communication. This would be true whether the scientist functioned as a solipsist or a materialist. The distinction would seem to constitute one of the distinguishing operations of empirical science. Thus, you may find discrepancies in my report so that it is plain that I did not report the revolutions of the kymograph correctly; or you may doubt that I could have watched the behavior of the rat and the kymograph accurately at the same time.

---

*Received in the Editorial Office on December 20, 1952, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

As a matter of practical operation, nothing may have been public about the whole process until I turned my data over to the typist. Or let us suppose, according to Boring's theory, that my discriminations are brain processes (2) and in his sense public; they still remain private in the necessary social sense of science, and under the most optimistic physiological research program will probably remain so for years to come. Their present public nature in Boring's sense is purely academic theory. Nor is it even necessary for behavioristic operations, as Watson recognized in admitting a sentient observer (17, p. 26f).

The argument really turns on what shall be taken as the observational point of reference behind which no questions will be immediately raised regarding the nature of the observational process. In introspectionism the report comes after the introspection; in behaviorism it comes after the "verbal report." Thus behaviorism and introspectionism differ in the basic scientific operation of distinguishing between observation and report, an epistemological operation. The fact that the report must be objective is irrelevant. To assume that the observation must be basically physical is to assume a materialistic metaphysics which is unnecessary in making the distinction. If one makes a solipsistic assumption, as Pratt (9) tentatively does, then the basically physical nature of the observation disappears.

B. The Operational Defining Technique

Although Kantor (8) frankly takes operationism as a postulational device in which he seems to accept a realism, others, he points out, take it as a defining technique. Thus, in attacking a proposition in connection with Kantor's realism, Stevens asks: "Can any operations be formulated which will either prove or disprove this proposition? If not, it is operationally meaningless, however much 'emotional meaning' it may pack" (13, p. 231f). He even finds it appropriate to give an operational definition of the observational experience itself of the scientist. He says that *"to experience* is, for the purpose of science, *to react discriminatively"* (12, p. 521). Yet Stevens argues that operationism is not a behaviorism: "Operationism does not deny images, for example, but asks: What is the operational definition of the term 'image'?" (13, p. 231). Again: "The division of concepts into the categories of subjective and objective is justifiable—if at all—only on pragmatic grounds, and only *provided* both types of concept answer the operational test" (*Ibid*).

Stevens believes, with Bills, that " 'mentalistic' concepts like percept, image, and idea can be operationally defined" (*Ibid*). But the mental concepts of Stevens are objective discriminations, indistinguishable from behavioristic ones. This result is brought about by a defining technique which is restricted to only certain kinds of processes. Stevens holds that images must be defined finally in terms observed by "the other one." There must be an independent experimenter, and the experimenter must be distinguished from the thing observed.

If a physicalistic terminology had not already practically done so, these rules of operating would logically lead the psychologist inevitably to a behaviorism, because the operational requirements set up preclude the use of such a distinction as Wundt's between mediate and immediate experience, by allowing only verbal report. Elsewhere Stevens writes: "Science demands public rather than private facts. 'Objectivity' in science is attained only when facts can be regarded as independent of the observer; for science deals only with those aspects of nature which all normal men can observe alike" (11, p. 327). In the last sentence of this quotation, the last clause is irrelevant since there has been historically demonstrated agreement among numerous scientists on immediate as well as mediate experience. The first clause in the sentence is critical to the argument for it excludes introspective observation. If the experimenter must have experience to be a scientist—and certainly many of the behaviorists from Watson on will admit this,—then he might have the experience in the rôle of an introspectionist as well as in the rôle of an objective observer. When the former is ruled out, an epistemological assumption is used in the definition of operationism which thereby necessitates the behaviorism arising from it.

C. INTRUSION OF ONTOLOGICAL ASSUMPTIONS

Stevens' epistemology really stems from a working ontological assumption for science in which only public and so objective reality can be the basis for empirical definition. "Only the reported part of 'experience' can ever get into science; the rest is meaningless since we have no operations for dealing with it" (12, p. 522).

Despite the fact that both Stevens and Kantor think they have avoided most metaphysical difficulties, they disagree on whether mental concepts can be operationally defined, and the disagreement rests largely on ontological assumptions, as these affect the operational definitions. Bills (1) asks appropriately for an operational definition of "operational." He points out that in the realism of Kantor, operational must mean "direct and immediate responses to objects and events," not to secondary constructs, as are light waves, which cannot therefore provide an operational basis for sensations.

On the other hand, operations are defined by Stevens in terms of the operation of definition, which is discrimination, and since "introspection is a discriminating operation," it can be used for defining concepts (p. 388). We should point out that both the introspection and the image or sensation so defined are objective discriminatory processes, because the mental processes in the subjective sense are not available in Stevens' ontology.

They do seem to be available, however, in the assumed metaphysics of Bills. Although he says that the dualism-monism dilemma is a meaningless problem in operationism, the "ready transposability of operationally defined concepts" produces in effect a monism. Parallelism becomes double-aspect monism "because where two sets of defining operations coincide perfectly they become identical operationally" (*Ibid,* p. 379). So mental concepts can be defined operationally and can be translated into either brain or behavior terms. The error in Bills' argument is that, even though physical operations may, all the necessary defining processes do not coincide perfectly because if one accepts only those philosophic assumptions which exclude introspection, one does not have the phenomena of awareness. If one does accept philosophic assumptions which permit introspection, then the usual variety of philosophic positions embracing these phenomena will be available. Thus differences among operational thinkers as to whether images can be operationally defined rest upon philosophic assumptions, including ontological ones.

D. The Variety of Behavioristic Concepts

Within the operational constructions of behavioristic experimentalists, other marked ontological differences are significantly apparent. The behaviorism of Watson was fundamentally, in the early stages at least, a methodological behaviorism, emphasizing the externally observable, what might be called natural phenomena. It was inevitable that externally observable stimulus and response would not always be adequate in an objective program of prediction and control, and that in the more mature stages of the program, intervening variables between stimulus and response would be constructed. Thus in the human field we have implicit (thinking) and the unverbalized (unconscious) responses of Watson (17, 18), and in animal behavior the symbolic (image-substitute) responses of Hunter (7). Watson and Hunter thought of these processes as naturalistically real ones—perhaps tangible is a better term,—real as a muscle twitch. Tolman calls the older behaviorism "physiological" and molecular, his own molar—a difference in level of explanation (15, p. 101ff). More importantly, however, in later behavioristic theory, many of the concepts of Tolman (15) and Hull (5), although still derived from or related to objective phenomena, have themselves no necessary, objectively observable, tangible reality, but may, as Hull states (6), forever remain logical constructs. In the earlier tradition, Solomon has indeed tried to give a tangible reality to the Hull concept of reaction inhibition (10, p. 4f). Now a world of symbolic constructs is certainly not a world of tangible, observable reality, so that working ontological differences exist in these scientific systems of concepts.

They must also differ in the epistemological devices used for bringing meaning to the objective, behavioristic observations. This difference is accentuated in the mathematico-deductive methodology of Hull (4, 5, 6), with its basic postulates, and hypotheses which are tested by seeking whether they are experimentally refuted. This philosophic rationalism stands in strange contrast to the empiricism of the classical behaviorists, with their inductive generalizations and limited hypotheses selected by experimental fact from alternative possibilities.

Moreover, in accepting means-end relationships as a basic fact of behavior, and as an intrinsic factor in the definition of even the facts of physics, Tolman is admittedly purposive (16), and clearly teleological in his scientific ontology. This organization of reality differs fundamentally from the ateleological ones sought by Hull (5) and the classical behaviorists (cf. 7, 17).

E. The Knowable Reality

One need not be a philosophic idealist to realize that the assumption of man's own awareness, images, and thoughts as constituting the only knowable reality would radically change such operational "systems" as those of Boring and Stevens, and would be most apt to do so at the point where operations themselves are defined. The introspective awareness would have a reality; tangible and concrete would perhaps mean only generally agreed upon, consistent, stable, and consequential; physical might have scientific meaning in psychology only in motivation as distinguishing the order of certain meaningful arrangements from others.

Since for Tolman immediate experience is the "only tangible real that we have" and the "common matrix out of which both physics and psychology are evolved" (15, p. 96), one is not surprised that Pratt is baffled at Tolman's behaviorism (9, 93ff). The paradox is underscored by Tolman's admission of the value of introspection in finding intervening variables in complex personality studies (15, p. 128). Once found, however, these variables are apparently no different from those derived from objective studies: mere symbols of objective means-end relationships. So far as his science is concerned, then, Tolman escapes solipsism by reliance on an instrumental pragmatism, in which the knowable mind is identified with functional relationships of objective adjustment, such that "the observed behavior itself turns out to be a set of operations performed by the observed organism relative to its own environment" (*Ibid,* p. 116). This ontological position is complemented by an epistemological one restricting observation to the objective type, or as he later says, to the positivistic "concrete repeatable operations" (p. 115).

Stevens tries to avoid metaphysics by assuming that the scientist and his world of observation are engaged in reaction, which is defined by other reaction, etc., in an infinite regress, which he refuses ultimate definition by plunging, as he says, *in medias res* (14). In so doing Stevens substitutes a world of pure reaction for the world of pure psychics of the solipsist, or for some mixed world of the realist or double-aspect monist. We wonder not that Stevens jumped, but what *res* he jumped into. However vague, his metaphysics is not solipsistic.

Bills agrees with Kantor that defining operations must be made upon "real concrete phenomena, not upon secondary constructs." But he cannot agree with Kantor "as to what constitutes reality or concreteness" (1, p. 388). Kantor insists "that discrimination must always be regarded as a response of the organism to colored or sounding objects and not a complex of 'experiences' and/or neural or general physiological action" (8, p. 40). In the same connection he states: "Probably a criterion for genuine operational data and procedures is whether or not one requires a double-aspect or identity dodge or organic bases or correlates to get rid (verbally) of dualism" (*Ibid*). If his naturalistic view is a realism, he says, "then realism or metaphysics is simply another name for science" (*Ibid,* p. 42). However much Bills and Kantor may be criticized for their ontological secularism, they at least both recognize an ontological function in the operational process. Kantor is correct in implying that a discrimination must always be between something (cf. 3, p. 468). The disagreement as to the nature of this something is a basic cause of differences.

For the empirical scientist, there is a reality of his observations; there is also a reality in the systematic formulations of his science. The nature of and relationships between these realities provide his ontology, which inextricably bound up with his epistemology, constitute the working philosophy of his science, which appears to be inescapable, even for operationists.

F. CONCLUSIONS

We see then that in scientific psychology epistemological and ontological assumptions are determining factors (*a*) in the definition of basic empirical observations, (*b*) in the rules whereby these observations are defined, and (*c*) in the types of constructs derived from them. This has been true even when in behavioristic experimentation or in explicit operationism, effort was made to escape metaphysical and epistemological determinants. Rather than operationists escaping philosophic questions, their adherence to particular epistemological and ontological assumptions, sometimes implicitly, accounts for fundamental differences among themselves, for differences between them and the introspectionists, and for the generally behavioristic character of their operational systems. These differences are especially significant in psychology because they include, in addition to differences with respect to type of acceptable explanation, as perhaps they do also in physics, differences as to the very existence at the phenomenal level of what is to be explained.

REFERENCES

1. BILLS, A. G. Changing views of psychology as science. *Psychol. Rev.*, 1938, **45**, 377-394.
2. BORING, E. G. A psychological function is the relation of successive differentiations of events in the organism. *Psychol. Rev.*, 1937, **44**, 445-461.
3. GRIFFITH, C. R. Principles of Systematic Psychology. Urbana: Univ. Illinois Press, 1943.
4. HULL, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, **44**, 1-32.
5. ————. Principles of Behavior. New York: Appleton-Century, 1943.
6. HULL, C. L., HOVLAND, C. I., ROSS, R. T., HALL, M., PERKINS, D. T., & FITCH, F. B. Mathematico-Deductive Theory of Rote Learning. New Haven: Yale Univ. Press, 1940.
7. HUNTER, W. S. Human Behavior. Chicago: Univ. Chicago Press, 1928.
8. KANTOR, J. R. Current trends in psychological theory. *Psychol. Bull.*, 1941, **38**, 29-65.
9. PRATT, C. C. The Logic of Modern Psychology. New York: Macmillan, 1948.
10. SOLOMON, R. L. The influence of work on behavior. *Psychol. Bull.*, 1948, **45**, 1-40.
11. STEVENS, S. S. The operational basis of psychology. *Amer. J. Psychol.*, 1935, **47**, 323-330.
12. ————. The operational definition of psychological concepts. *Psychol. Rev.*, 1935, **42**, 517-527.
13. ————. Psychology and the science of science. *Psychol. Bull.*, 1939, **36**, 221-263.
14. ————. Psychology: The propaedeutic science. *Phil. of Sci.*, 1936, **3**, 90-103.

15. Tolman, E. C. Collected Papers in Psychology. Berkeley: Univ. Calif. Press. 1951.
16. ———. Purposive Behavior in Animals and Men. New York: Appleton-Century, 1932.
17. Watson, J. B. Behavior. New York: Holt, 1914.
18. ———. Behaviorism. New York: Norton, 1925.

Department of Psychology
University of Kentucky
Lexington, Kentucky

(Founded in 1935 by Carl Murchison)

The Journal of PSYCHOLOGY

The General Field of Psychology

EDITED BY

CARL MURCHISON

Gordon Allport
Harvard University

Arthur G. Bills
University of Cincinnati

E. A. Bott
University of Toronto

Leonard Carmichael
Smithsonian Institution

Herbert S. Conrad
U. S. Office of Education

Elmer K. Culler
University of Rochester

John F. Dashiell
University of North Carolina

Roy M. Dorcus
University of California at Los Angeles

Horace B. English
Ohio State University

Franklin Fearing
University of California at Los Angeles

Frank A. Geldard
University of Virginia

Clarence H. Graham
Columbia University

Ernest R. Hilgard
Stanford University

H. M. Johnson
Tulane University

Heinrich Klüver
University of Chicago

Carney Landis
Columbia University

K. S. Lashley
Harvard University

Norman L. Munn
Bowdoin College

John Paul Nafe
Washington University

Donald G. Paterson
University of Minnesota

Lee Edward Travis
University of Southern California

Robert C. Tryon
University of California

Michael J. Zigler
Wellesley College

If this space should be unstamped, this is the regular library edition. But if this space is stamped with a designating title, this is a special edition, sold under the restrictions of a bilateral contract, and may not be resold for a period of five years from date of publication.

VOLUME 36

1953

Copyright, 1953, by The Journal Press
Published Quarterly by the Journal Press
Provincetown, Massachusetts, U. S. A.

AUTHOR INDEX

Auble, Donavon . . . 165, 187
Ayad, Joseph-Magdy . . . 295
Barratt, Ernest S. . . . 17
Bartley, S. Howard . . . 201, 401, 409
Bechtoldt, Harold P. . . . 83, 111
Bendig, A. W. . . . 427
Berg, Irwin A. . . . 3
Bousfield, W. A. . . . 67
Brody, David S. . . . 347
Carpenter, Lewis G., Jr. . . . 131
Carter, Darrell B. . . . 299
Cohen, B. H. . . . 67
Crannell, C. W. . . . 417
Cumings, Ruth . . . 195
Edgington, Eugene S. . . . 51
Farber, Maurice L. . . . 243
Farnsworth, Paul R. . . . 295
Fattu, Nicholas A. . . . 153
Foltin, Edgar M. . . . 251
Gengerelli, J. A. . . . 127
Gordon, Leonard V. . . . 387
Hammer, Emanuel F. . . . 357
Harris, Albert J. . . . 323
Hauty, G. T. . . . 143, 363
Hochberg, Carol Barnes . . . 341
Hochberg, Julian E. . . . 341
Jones, F. Nowell . . . 207
Jones, Margaret Hubbard . . . 207
Kline, Milton V. . . . 195
Knehr, Charles A. . . . 289
Levitt, Eugene E. . . . 393
Lewis, Don . . . 83, 111
London, Ivan D. . . . 379
McAllister, Dorothy E. . . . 83, 111
Mech, E. Victor . . . 153, 165, 187
Michael, John L. . . . 127
Middleton, Charles E. . . . 423
Nissen, Henry W. . . . 271
Ojemann, Ralph H. . . . 393
Politz, Alfred . . . 11
Roff, Merrill . . . 347
Roswell, Florence G. . . . 323
Schaerer, Robert W. . . . 187
Schlosberg, Harold . . . 435
Schofield, William . . . 59, 137
Sheehan, Joseph . . . 27
Stanley, Walter C. . . . 435
Stein, Morris I. . . . 311
Sumner, F. C. . . . 423
Wendt, G. R. . . . 143
Wilkinson, Frank R. . . . 201

TABLE OF CONTENTS

The reliability of extreme position response sets in two tests 3
By **Irwin A. Berg**

Questionnaire validity through the opinion-forming question 11
By **Alfred Politz**

An analysis of verbal reports of solving spatial problems as an aid in defining spatial factors 17
By **Ernest S. Barratt**

Theory and treatment of stuttering as an approach-avoidance conflict . . 27
By **Joseph Sheehan**

Kinesthetically guided movements of head and arm 51
By **Eugene S. Edgington**

A study of medical students with the *MMPI*: I. Scale norms and profile patterns 59
By **William Schofield**

The effects of reinforcement on the occurrence of clustering in the recall of randomly arranged associates 67
By **W. A. Bousfield and B. H. Cohen**

Correlational analysis of the learning and relearning of four different tasks on the modified Mashburn apparatus 83
By **Don Lewis, Dorothy E. McAllister, and Harold P. Bechtoldt**

Correlational study of performance during successive phases of practice on the standard and reversed tasks on the *SAM* complex coördinator . . 111
By **Don Lewis, Dorothy E. McAllister, and Harold P. Bechtoldt**

A non-parametric test for the reliability of the difference between two proportions 127
By **J. A. Gengerelli and John L. Michael**

An experimental test of an hypothesis for predicting outcome with electroshock therapy 131
By **Lewis G. Carpenter, Jr.**

A study of medical students with the *MMPI*: II. Group and individual changes after two years 137
By **William Schofield**

Studies of vestibular function: I. The duration of primary nystagmus as a function of speed of rotation and of acceleration 143
By **G. T. Hauty and G. R. Wendt**

Interruption: Its effect upon performance in a "troubleshooting" situation 153
By **Nicholas A. Fattu and E. Victor Mech**

Partial verbal reinforcement related to distributed practice in a classroom situation 165
By **Donavon Auble and E. Victor Mech**

The effects of "set" on group performance under strong auditory stimuli 187
By **E. Victor Mech, Robert W. Schaerer, and Donavon Auble**

A study of learning characteristics of public health nurses in relation to mental health education 195
By **Ruth Cumings and Milton V. Kline**

Some factors in the production of gamma movement 201
By **S. Howard Bartley and Frank R. Wilkinson**

Modern theories of olfaction: A critical review 207
By F. Nowell Jones and Margaret Hubbard Jones

English and Americans: Values in the socialization process 243
By Maurice L. Farber

Personality traits of psychotic patients as revealed in their spontaneous paintings 251
By Edgar M. Foltin

Sensory patterning versus central organization 271
By Henry W. Nissen

Individual differences in subjective color 289
By Charles A. Knehr

Shifts in the values of opinion items: Further data 295
By Joseph-Magdy Ayad and Paul R. Farnsworth

A further demonstration of phi movement cerebral dominance 299
By Darrell B. Carter

Creativity and culture 311
By Morris I. Stein

Clinical diagnosis of reading disability 323
By Albert J. Harris and Florence G. Roswell

Familiar size and subception in perceived depth 341
By Carol Barnes Hochberg and Julian E. Hochberg

Appearance and choice status during adolescence 347
By Merrill Roff and David S. Brody

The possible effects of projective testing upon overt behavior 357
By Emanuel F. Hammer

Response similarity-dissimilarity and differential motor transfer effect . . 363
By George T. Hauty

An application of the engrammatic concept 379
By Ivan D. London

A factor analysis of the 48 Szondi pictures 387
By Leonard V. Gordon

The aims of preventive psychiatry and "causality" as a personality pattern . 393
By Eugene E. Levitt and Ralph H. Ojemann

The perception of size or distance based on tactile and kinesthetic data . . 401
By S. Howard Bartley

Conditions intended to manipulate accommodation as factors in performance with a visual acuity target 409
By S. Howard Bartley

A preliminary attempt to identify the factors in student-instructor evaluation . 417
By C. W. Crannell

Season of birth as related to seasonal preference and personality traits . . 423
By Charles E. Middleton and F. C. Sumner

Student achievement in introductory psychology and student ratings of the competence and empathy of their instructors 427
By A. W. Bendig

The psychophysiological effects of tea 435
By Walter C. Stanley and Harold Schlosberg

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 3-9.

THE RELIABILITY OF EXTREME POSITION RESPONSE SETS IN TWO TESTS*

Department of Psychology, Northwestern University

IRWIN A. BERG[1]

A. THE PROBLEM

There is ample evidence that responses to test items are sometimes highly individual expressions of "set." Cronbach (3, 4) has described a number of these response sets, such as *acquiescence* (tendency to say "true," "yes," "agree," etc.), tendencies to work for *speed* rather than *accuracy*, tendencies to guess, etc. He notes that when items are difficult or ambiguous, the influence of response sets is most marked.

The indications are that at least some forms of response sets are reflections of personality traits. Lorge (5), for example, reported that *confidence* scores on the *Bernreuter Personality Inventory* correlated .27 with tendencies to circle "yes" on several tests. Berg and Collier (1) found that groups of persons who were operationally defined as maladjusted made significantly more extreme position choices than normal subjects on the *Perceptual Reaction Test*. While the available evidence indicates that response sets are reasonably stable for a given test, an unanswered question concerns the degree of response set consistency in different test situations. As Cronbach (3, p. 486) succinctly put it: "The crucial question for an understanding of response sets is the extent to which they are transient or fixed."

Rundquist (6) studied consistency in taking extreme positions on a test of personality and a test of interest which were administered in immediate succession to 111 factory girls. He found the tendency to make extreme choices in a 5-alternative test situation correlated .40 for the two series. He concluded that there was a real tendency for those subjects who took extreme positions in describing their personality traits to take extreme positions also when describing their interests. However this tendency, he believed, was not sufficiently pronounced to make it a useful measure for predicting a criterion such as behavior on a job.

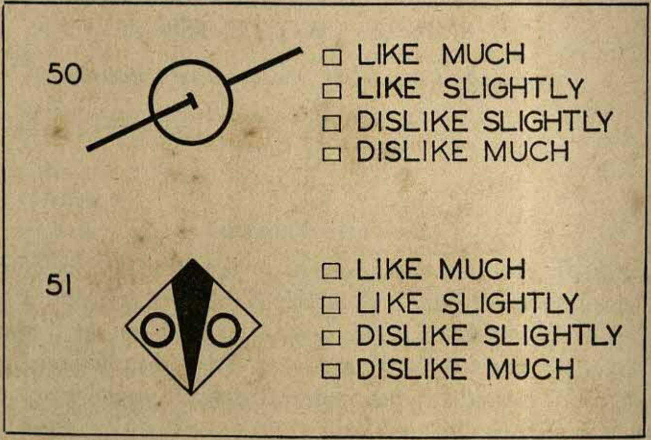
B. PURPOSE

The present study aims at exploring extreme position response sets over two intervals of time and with two tests which use qualitatively different items.

---

*Received in the Editorial Office on January 2, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to express his appreciation to Mrs. Joanne S. Collier, who rendered valuable statistical assistance in the present study.



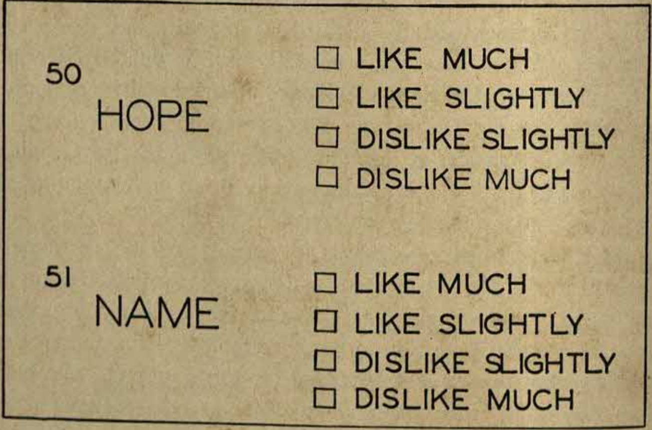
50

☐ LIKE MUCH
☐ LIKE SLIGHTLY
☐ DISLIKE SLIGHTLY
☐ DISLIKE MUCH

51

☐ LIKE MUCH
☐ LIKE SLIGHTLY
☐ DISLIKE SLIGHTLY
☐ DISLIKE MUCH

FIGURE 1
SAMPLE ITEMS FROM THE *Perceptual Reaction Test*



50 HOPE

☐ LIKE MUCH
☐ LIKE SLIGHTLY
☐ DISLIKE SLIGHTLY
☐ DISLIKE MUCH

51 NAME

☐ LIKE MUCH
☐ LIKE SLIGHTLY
☐ DISLIKE SLIGHTLY
☐ DISLIKE MUCH

FIGURE 2
SAMPLE ITEMS FROM THE *Word Reaction Test*

The tests used were the *Perceptual Reaction Test* (2), hereafter referred to as *PRT*, which is composed of 60 abstract designs and the *Word Reaction Test*, hereafter referred to as *WRT*, which was specially prepared for the present study by taking 60 words from the Wells-Ruesch word association list (7, p. 97) and presenting them in the same form as the *PRT* (see Figures 1 and 2 for samples of the items). To each of the 60 items in either test, the subject was required to respond by marking one of the following alternatives: *like much, like slightly, dislike slightly,* or *dislike much.* The score was the number of times the extreme position, either *like much* or *dislike much*, was chosen by the subject.

C. Subjects and Procedure

Two groups of subjects were used: a group of 95 students in elementary psychology, retested after an interval of 15 days, and a group of 22 students in an undergraduate course in industrial psychology, retested after an interval of 7 days. For the initial testing the *PRT* was given first, followed immediately by the *WRT*. The order was then reversed for the retest.

Both tests were equally unstructured or ambiguous in the sense that there were no "right" or "wrong" answers. Also, both tests used identical 4-alternative choice situations for each item. Thus it was presumed that both tests provided approximately equal opportunity for extreme response sets to appear. However, while both tests used items of visual form, it is believed that the tests may be regarded as qualitatively different since the *PRT* used only nonverbal items and the *WRT* used words. This arrangement permitted cross-comparisons to be made between extreme response scores for two qualitatively different tests and for retesting at different time intervals.

D. Results and Discussion

As reference to Table 1 will indicate, the similarity of the mean scores for both groups in the present study and for both test-retest intervals affords some evidence of extreme position response stability insofar as central tendencies are concerned. More pertinent than such group tendencies, however, is the reliability of the extreme position choice over a period of time. Of course, if a particular response set validly reflected a transient personality facet such as mood, it might conceivably have some slight usefulness as a measure of that facet, even though the set were as unstable in time as the condition it measured. But the chief value of response sets as possible measures of personality or other characteristics will be in terms of their power to reflect validly basic personality dimensions. For this a high degree of stability in time is essential.

The data in the present study indicate that one response set form, the tendency to choose extreme positions when marking choices on the *PRT* and *WRT*, is acceptably stable over test-retest periods up to 15 days. For 7 and 15 day intervals, as shown in Table 3, the *PRT* reliability coefficients are respectively .85 and .78 while *WRT* reliabilities are .84 and .94 for the same intervals. This level of reliability is sufficiently high to encourage the use of extreme position response sets as measures of personality, provided the existing validity evidence of the relationship of such sets to personality is considerably extended and confirmed.

TABLE 1

EXTREME RESPONSE MEAN SCORES FOR TWO GROUPS AND TWO TEST-RETEST INTERVALS

Test*	N	Mean	*SD*
PRT-1	95	18.5	9.4
PRT-1	22	20.0	10.0
PRT-2	95	19.1	10.3
PRT-2	22	21.0	10.5
WRT-1	95	21.8	11.3
WRT-1	22	25.4	12.7
WRT-2	95	21.4	13.9
WRT-2	22	24.5	10.2

*The numbers 1 and 2 in PRT-1, WRT-2, etc., refer to the first and second test administration. For the group of 95 the test-retest interval was 15 days and for the group of 22 it was 7 days.

Since extreme position response sets are stable for the *PRT* and the *WRT* individually, it is reasonable to hypothesize that similar consistency should be manifested from one to the other since both tests are presumably unstructured to the same degree. As a test of this hypothesis, the *PRT* and *WRT* were given in immediate succession to two groups of 22 and 95 subjects each. The same tests were again administered in immediate succession to the same groups after an interval of one to two weeks. The results of this test-retest comparison are presented in Tables 2 and 3. For the *PRT* and *WRT* administered in immediate succession the correlations from one test to the other range from .34 to .77 while the range is .30 to .64 for the 15 and 7 day intervals respectively. The general trend may be said to be in the neighborhood of .40, about the same degree of relationship found by Rundquist (6) in his study of extreme position responses for tests of personality and interest.

TABLE 2

EXTREME RESPONSE SCORE CORRELATIONS FOR THE *PRT* AND *WRT* ADMINISTERED IN IMMEDIATE SUCCESSION

Tests	N	r
	22	.53
PRT-1 and WRT-1	95	.41
PRT-2 and WRT-2	22	.77
	95	.34

It seems clear that the tendency to adopt extreme response positions on the *PRT* carries over in real fashion to the *WRT*, but the relationship is only moderate. Rundquist suggested that the level of consistency with which his subjects carried extreme response sets from one test to another was due to mood or some other temporary condition, rather than to anything basic about the individual. However, Rundquist did not retest his subjects after an interval of time; hence he had no data on the assumed temporary nature of such sets. In the present study the *PRT-WRT* correlations, though modest in size, are all significantly different from zero even after a retest interval of 15 days. Thus it seems probable that more than a transient state of the subject is being reflected by extreme response sets.

TABLE 3

TEST-RETEST RELIABILITY OF *PRT* AND *WRT* EXTREME RESPONSE SCORES FOR TWO TIME INTERVALS

Tests	7-day Interval N = 22 r	15-day Interval N = 95 r
PRT-1 & PRT-2	.85	.78
WRT-1 & WRT-2	.84	.94
PRT-1 & WRT-2	.64	.38
WRT-1 & PRT-2	.56	.30

Originally, it had been assumed that the *PRT* and *WRT* were equally unstructured. Since response sets are called forth to the degree to which a test lacks structure, the tendency to adopt extreme position responses on the one test should carry over markedly to the other. The fact that this carryover was moderate rather than marked raises the question of whether the tests were really equally unstructured. Oral reports from 20 of the 117 subjects used indicate that the tests were probably not equal in this respect. The *WRT* with its verbal content seemed to arouse highly specific and enduring affective associations[2] to a much greater extent than the *PRT*. If such were generally the case, the test-retest reliability of the *WRT* should be expected to be higher than that of the *PRT* because of the influence of memory. Since data in the present and other studies indicate that the *WRT* retest reliability is usually higher than the *PRT*, some indirect support is lent to the hindsight observation that the tests were unequally structured.

---

[2] For example, a number of subjects reported that they checked "like much" for the word *hope* because "hope is what makes life worth living" etc. Reference to their test responses indicated that the subjects had correctly recalled how they had expressed their reactions. Relatively few *PRT* designs aroused such specific associations.

The fact that a response set as studied in the present investigation carries over only moderately from one test to another does not, of course, limit the potential usefulness of the set as a possible measure of personality or other traits when only a particular test is involved. It does mean that a given response set may be found to reflect validly some dimension as emotionality on one test yet bear little relationship to emotionality on another, different test. In consequence, the future value of response sets as measures of personality characteristics will be limited to specific tests until a method of equating structure in various testing situations is developed.

Most studies thus far have given primary consideration to ways of controlling or eliminating response sets. Since response sets affect test reliability and validity, such studies are extremely important. However, the sets themselves appear to be quite stable, and there is evidence, though slender and as yet inconclusive, that certain response tendencies reflect basic personality traits. Accordingly, it would seem profitable to use the sets rather than eliminate them, at least for certain testing situations. With some inspired research and grubbing analysis, it should eventually be possible to develop a test, as of mathematics, which could be scored for mathematics proficiency on some items and for personality dimensions by means of response sets on other items. With such multiple scoring in mind, it is conceivable that items designed to arouse certain response sets could be deliberately inserted in future tests.

E. Summary

Extreme position response set scores were obtained for groups of 22 and 95 subjects by counting the number of times an extreme position (i.e., checking either *like much* or *dislike much*) was adopted when responding to items with four options in a verbal and a nonverbal test. The tests were administered in immediate succession to each group. The same tests were again administered to the group of 22 subjects after an interval of 7 days and to the group of 95 subjects after 15 days. Extreme position response sets were found to be stable for a given test. Test-retest reliability coefficients ranged from .78 to .94. The tendency to mark extreme position choices on one test carried over moderately to the other test. Extreme position scores on the verbal test correlated from .34 to .77 with similar scores on the nonverbal test for the two groups in two test situations. It is believed that such extreme response sets are sufficiently stable in certain tests to merit careful study as possible measures of personality or other characteristics. While response sets have potential usefulness as measures of various traits, their value will be limited to a specific test until a method for equating different tests for structurization is developed.

REFERENCES

1. BERG, I. A., & COLLIER, J. S. Personality and group differences in extreme response sets. *Educ. Psychol. Measmt.,* (accepted for 1953 publication).
2. BERG, I. A., & HUNT, W. A. The Perceptual Reaction Test. Evanston, Illinois, 1949.
3. CRONBACH, L. J. Response sets and test validity. *Educ. Psychol. Measmt.,* 1946, **6**, 475-494.
4. ———. Further evidence on response sets and test designs. *Educ. Psychol. Measmt.,* 1950, **10**, 3-31.
5. LORGE, I. Gen-like: halo or reality. *Psychol. Bull.,* 1937, **34**, 545-546.
6. RUNDQUIST, E. A. Response sets: a note on consistency in taking extreme positions. *Educ. Psychol. Measmt.,* 1950, **10**, 97-99.
7. WELLS, F. L., & RUESCH, J. *Mental Examiners' Handbook* (second edition). New York: Psychological Corporation, 1945.

Department of Psychology
Northwestern University
Evanston, Illinois

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 11-15.

QUESTIONNAIRE VALIDITY THROUGH THE OPINION-FORMING QUESTION*

Alfred Politz Research, Inc.

ALFRED POLITZ[1]

The opinion questionnaire is one of the most useful of research instruments in social psychology, and it is, of course, fundamental to market research. In the chips-down situation of market research the questionnaire must predict. If *action* does not follow on opinion, market research comes swiftly to an end. Our concern about questionnaire validity is therefore doubly motivated—both by scientific ideals and the urgent necessities of practice.

A lay audience can be entertained by a poll which asks "interesting" questions and obtains "interesting" answers. In this area, it is often assumed that the merit of a survey is demonstrated by the fact that the questions were unbiased.

If research is to be converted from more or less "interesting" descriptions to a predictive instrument, the unbiased question loses its value. In fact, the biased question becomes an essential instrument of investigation. Since the "correct" answers are not necessarily the ones that pop up when an unbiased question is asked, the biased question, or a set of biased questions, will frequently furnish the only solution. This situation is by no means unique to opinion polling. In the case of court trials jurists have never accepted the naïve assumption that truth can be obtained from the defendant by an unbiased question. In fact, the court is split into two camps, with its members firing barrages of highly biased suggestions and leading questions at the defendant. It is under the impact of the most violent biases contained in these questions and suggestions that the truth comes out.

Overlapping the problem of biased questions is another issue. How useful are open-end questions as compared with a choice of propositions or suggestions from which the respondent has to select what comes closest to his opinion? If the peculiar selectivity of superlatives is properly discounted, choice of suggestions is an essential tool in productive research. The disadvantage is the enormous amount of work that usually has to go into a correct definition of alternatives.[2] In contrast, the open-end question seems the climax of simplicity for the questionnaire designer. It does not require imagination or structural knowledge.

---

*Received in the Editorial Office on January 2, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] I wish to thank Dr. John Dollard, Professor of Psychology at Yale University, for having read this paper and for making some useful suggestions in regard to it.

These easy questions look enticing to the superficial observer. Seemingly they provide for the respondent the full latitude of freedom. Open-end questions are frequently defended as being unbiased, since no influencing thought is evoked in the respondent by the interviewer. The danger of bias is, however, not removed. It is only *transferred* from the researcher to the respondent. The answers given by the respondent to an open-end question express those thoughts that can be put into words with the least amount of work and embarrassment; but they are often not a basis for predicting action. After all, why should the casual respondent bear the anguish of making a choice or forming a novel opinion?

Behind the enticingly simple open-end questions stands the illusion that the public can and will do the work for the researcher. "Why do you dislike Candidate X?" "What are the advantages of Product X?" "What improvements do you think the manufacturer should make?" are questions that produce answers, and thereby statistical reports, but they do not necessarily produce *valid* answers.

In the year 1800, most people lighted their rooms with kerosene lamps. Let us imagine that an opinion survey had been undertaken at that time with the following unbiased and carefully worded question: "Suppose money were no object. How do you think the lighting system in your home could be improved?" The respondents who had seen nothing but kerosene lamps and candles could hardly think of a lighting system where the lamp was mounted on the ceiling and the light could be switched on and off by means of a little gadget located within easy reach beside the entrance. If the answers of lamp consumers in the year 1800 had been used to guide the lamp manufacturers, we would be lighting our rooms nowadays with vastly improved kerosene lamps. Mere lack of bias in the question does not guarantee a useful answer.

Mass questioning which assumes that people have opinions and that consumers know what they want is the opposite of predictive research.

It cannot be a surprise that in predictive research the operational rules are similar to those in physical sciences, where the relationship between experimental design and prediction has been understood for more than three centuries (1). Answers to questions can no longer be considered solutions to problems. Answers are just responses to specific stimuli. Control and definition of the stimulus determines the interpretation of the response. Seen in this light, the question of whether the response is biased or unbiased is no longer relevant. Responses are reactions observed under specifiable conditions. The "conditions" are the environment at the time of the interview, the appearance of the interviewer, the internal state of the subject, and the questions asked.

---

[2] Note, for instance, the labor which S. A. Stouffer and his associates have put into making sure that opinions are scalable (5).

If a "biased" question is needed to get a true answer, and if a "biased" answer is needed to serve as the basis of a prediction, we are obtaining valid research.

Within this larger frame of reference, the opinion-response is seen as one factor from which the conclusions are to be drawn. The correctness or incorrectness of the opinion, the question of whether it is intrinsic or natural with the respondent or influenced by the interviewing procedure ceases to have bearing on the purpose of the polling operation. The chemist who mixes the ingredients in certain ratios or subjects them to certain types of stimuli such as increase or decrease of temperature, the presence of a magnetic field, or the addition of catalysts, is aware of the fact that he influences what is happening within the test tube, but he has to do it in order to determine what will happen in times and places other than those at hand. He wants to predict. He does not hope to obtain the solutions to his problems from the thinking of the chemical ingredients. He does not depend on their opinions. Neither does the predictive research depend solely on the opinions of the respondents. The various opinions are formed during the various stages of the interview and are to be seen as characteristics of individuals under specific conditions.

Indeed if a prediction of people's behavior is the purpose of research, the assumption of "uninfluenced" opinion may be profoundly misleading. The opinion ceases to be the kind of solid that conventional usage depicts it to be. An opinion is only the response to stimuli at a given level of knowledge. A change in the questions changes the opinion. A change in information changes the opinion. It is justifiable, in many situations, to say that people change their knowledge instead of complaining about the degree to which they are apt to change their opinions. The frequently-referred-to fickleness of the public in regard to opinions often reduces itself to a simple increase of knowledge on the part of the public.

For example, if someone has the "opinion" that taxes should be reduced, he might change this opinion once he hears that the tax money is needed for adequate defense. His knowledge level has changed. His opinion follows.

He may hear in the next step that adequate defense can be built up with taxes as they are if only waste and inefficiency can be avoided. His new opinion is that taxes should be kept at current levels.

He later finds himself confronted with a further increase of data. Authorities convince him that there is no need for defense at all. He now has the opinion that taxes should be reduced and if asked, "Do you think that taxes should be reduced?" he will respond to this stimulus and answer "yes."

If he is further asked if *sales taxes* should be reduced, the question furnishes knowledge which he has not had so far. The question implies that there are different types of taxes. This possibility had not occurred to the respondent before. The respondent now becomes uncertain as to whether the picture he had of the tax problem is good enough for having an opinion on it at all. Under the impact of the sales tax question, he forms a new opinion—i.e., "I should look more deeply into the problem as to what kind of taxes there are, how they are collected, and what the different parts are used for."

The increase in knowledge need not be brought from the outside. Suppose we ask a respondent, "Did you ride on a bus a year ago on September 18?" It is a case of knowledge related to an objective situation. At the time when the question is asked, the knowledge does not exist. However, the stimulus of the question induces the respondent to explore his memory, apply a reasoning procedure, and produce the new piece of knowledge in terms of a conclusion. His starting point is the memory that during the year the question refers to, he went to work every day by bus, except on holidays and during his vacation. The date mentioned can be identified as falling on a weekday outside of vacation time. Consequently, the respondent must have ridden on a bus.

The extent to which questions form knowledge and opinion is a matter of degree. If the stimulus of the question is nearly identical with stimuli to which the respondent has learned a response in the past, then the question stimulus will evoke the learned response, and almost no new opinion or new knowledge is created.[3] However, the degree of similarity of the new stimulus pattern to the older one cannot easily be determined during the interview.

If interviews could be conducted by free association as Freud suggests, then the old stimuli could be better identified. But interviews conducted in the opinion-attitude fields cannot resemble a psychoanalytical situation irrespective of how deeply the respondent is investigated (3).

The few relevant memories surrounding the stimuli under study are interwoven with a maze of other memories which cannot be readily segregated by the respondent. A large portion of the memories is uncomfortable. Before a respondent is willing to expose himself to discomfort, he must be motivated and must expect a reward.

The neurotic who submits to psychoanalytical treatment is impelled by neurotic misery and expects health as a reward (2). This powerful incentive is just strong enough to offset the powers of resistance to a productive interview. Through psychoanalysis a better insight into old stimuli is obtained. It goes without saying that this procedure too is opinion forming.

---

[3] A resumé of the experimental data on stimulus generalization has been made by Hovland (4).

The dependability of a market prediction increases to exactly the degree that all the conditions eliciting the opinion are understood and controlled. The ideal interview would present exactly the stimuli which would exist at the time, and under the circumstances, for which a prediction is sought.

For example, for almost a decade, the question "What brand of refrigerator will you buy next?" produced more votes for General Electric than for Frigidaire, although Frigidaire went on selling more.

The potential consumers whose answers produced such a piece of incorrect statistics may have expressed what they truly believed. However, when they finally approached the time for buying the new refrigerator, their "curiosity" made them look at advertisements of various refrigerators, made them read sales literature, made them listen to the comments of friends, made them look at the refrigerators of different manufacturers, made them study the shelves, the butter compartment, and listen to the salesman's arguments. In other words, the potential customer was exposed to opinion-forming stimuli altogether different from the stimuli of the interview. The latter would have produced a more useful response if the stimuli had closely resembled those surrounding the act of purchase. In other words, only an opinion-forming interview could have produced a correct prediction. The ideal interview, which provides stimuli identical with the situation for which a prediction is undertaken, does not exist in reality. However, the real stimuli serve as a model which the experimentor tries to approach. If the similarity is insufficient, the poll is useless.

Practical polling shows instances of both opinion-influencing interviews and attempts to avoid influence. Instead of vacillating in a theoretically untenable position it is better, first, to recognize the inevitability of influence and, then, to use it for predictive research in social psychology, advertising, and marketing.

REFERENCES

1. Cohen, M. R., & Nagel, E. An Introduction to Logic and Scientific Method. New York: Harcourt, Brace, 1934. (Chapter XX, pp. 391-403.)
2. Dollard, J., & Miller, N. E. Personality and Psychotherapy. New York: McGraw-Hill, 1950. (P. 13.)
3. Freud, S. A General Introduction to Psychoanalysis. New York: Liveright, 1920. (P. 96.)
4. Hovland, C. I. Human learning and retention. In *Handbook of Experimental Psychology* (S. S. Stevens, *Ed.*). New York: Wiley, 1951. (Pp. 616-617.)
5. Stouffer, S. A., Guttman, L., Suchman, E. A., Lazarsfeld, P. A., Star, S. A., & Clausen, J. A. Measurement and Prediction; Studies in Social Psychology in World War II, Vol. 4. Princeton, N. J.: Princeton Univ. Press, 1950. (Pp. 60-90.)

Alfred Politz Research, Inc.
270 Park Avenue
New York City 17

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 17-25.

AN ANALYSIS OF VERBAL REPORTS OF SOLVING SPATIAL PROBLEMS AS AN AID IN DEFINING SPATIAL FACTORS*

Department of Psychology, University of Delaware

ERNEST S. BARRATT[1]

A. PURPOSE

The present investigation was designed primarily to determine if a systematic analysis of introspective data could be of value in defining psychological processes used by subjects in solving problems on paper-and-pencil "space" tests. The general thesis was that one of the first steps in predicting "what" is being measured by tests of ability should be an analysis of the problem-solving processes used by subjects when taking the tests, especially when those tests are used as a basis for defining factors.

In this study a factor analysis of space tests was related to an analysis of the subjects' problem-solving behavior on the tests in order to arrive at an operational definition of three spatial factors.

A review of the literature concerned with the space factor indicates that: a space factor was one of the first factors isolated (5); it was the second most frequently isolated factor up to 1940 (9); recent research indicates that it may involve at least three types of abilities (4, 6); and factor analysis alone has not enabled the investigators to reach agreement on precise operational definitions of these space factors.

The three space factors being considered here were isolated by the Air Force psychologists during World War II (4) and also by Thurstone (6). It was hypothesized in the present investigation that these factors represented the following abilities: the first space factor represents the ability to rotate mentally figures or objects as presented in paper-and-pencil tests to determine whether the objects are alike; the second space factor represents the mental ability to determine what a given pattern or configurational relationship would be if the pattern or configurational relationship were changed so that its parts occupied a different physical relationship to each other in space; the third space factor represents the mental ability to reorient oneself spatially with relation to an object or

---

*Received in the Editorial Office on January 6, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Portions of this article were presented by the author at the American Psychological Association's 1952 meeting in Washington, D. C.

scene presented in a drawing and to determine either from what direction one is looking at the object or how other objects would appear if one were looking at them from the position of a given object. The slight changes made in these definitions in relation to those hypothesized by the Air Force psychologists and by Thurstone were made on the basis of the pilot study conducted by the author in this area.

B. Procedure

The procedure for the investigation was divided into two main parts:

1. The testing period, during which a battery of 10 paper-and-pencil tests were administered. The 10 tests used in this investigation were as follows. Tests predicted to measure Space 1 ability: (*a*) Thurstone's Flag Test (published by Science Research Associates); (*b*) Thurstone's Figures Test (published by Science Research Associates); (*c*) Thurstone's Card Test (published by Science Research Associates).

Test predicted to measure Space 2 ability:[2] (*d*) *DAT* Space Relations Test (published by Psychological Corporation).

Tests predicted to measure Space 3 ability: (*e*) Guilford-Zimmerman Aptitude Survey, Part V, Spatial Orientation (published by Sheridan Supply Co.); (*f*) Barratt-Fruchter Chair-Window Test,[3] an experimental test which is in the process of being developed; it was designed to measure the type of ability which the third space factor was hypothesized to represent (corrected split-half reliability coefficients have ranged from .89 to .94); (*g*) Industrial Aptitude Survey, Part 10, Spatial Orientation, Form *B* (published by the Psychological Research Center).

Test variables 8-10 were the linguistic portions of the *ACE* Psychological Examination for College Students, 1939 edition.

2. The second main part of the procedure was the personal interview period divided into three intervals: (*a*) the interval during which rapport was established with the subject and the subject was given instructions as to the procedure which would be followed in the interview; (*b*) the interval during which the subject's problem-solving processes on problems typical of the various tests were recorded verbatim, and (*c*) the interval during which the subject answered specific questions regarding his problem-solving procedures.

The interview periods ranged in time from approximately 50 minutes to one hour.

---

[2] Two other space two tests were scheduled to be given in the experiment, but they were not received in time to be administered with the remainder of the battery.

[3] Designed by E. S. Barratt, the University of Delaware, and Benjamin Fruchter, the University of Texas.

The use of verbal reports in analyzing thought processes has been frequent and varied (1, 3, 7, 8). In using this approach in the present experiment, the investigator assumed that the subjects were aware of some of the mental processes they used to solve the problems and that they were able to communicate these thoughts to the investigator.

C. SUBJECTS

The subjects were 84 male students at St. Edward's University, Austin, Texas. This group constituted the entire summer enrollment at the University except for approximately 10 students who could not participate in the experiment because it conflicted with their class schedules. The academic classification of the subjects was as follows: 21 freshmen, 20 sophomores, 23 juniors, and 20 seniors.

D. TREATMENT OF DATA AND RESULTS

The test intercorrelation matrix (Table 1) was factored using a computational method based on Thurstone's centroid method of factor analysis. Four factors were isolated which, after five orthogonal rotations, were identified as the three space factors and a verbal factor (Tables 2 and 3). All of the tests except one (the Guilford-Zimmerman Spatial Orientation Test) appeared on the specific factors in accordance with the investigator's predictions.

TABLE 1

MATRIX CONTAINING THE TEST INTER-CORRELATIONS (R MATRIX) (DECIMAL POINTS OMITTED. ALL ENTRIES ARE TO THREE DECIMAL PLACES.)

		Test									
Test		1	2	3	4	5	6	7	8	9	10
Flags	1		839	634	429	460	561	456	225	306	432
Figures	2	839		653	510	391	441	361	207	350	483
Cards	3	634	653		459	535	566	471	314	444	476
DAT Space Relations	4	429	510	459		586	591	453	272	352	524
Chair-Window	5	460	391	535	586		536	724	424	580	496
Guilford-Zimmerman Spatial Orientation	6	561	441	566	591	536		661	362	506	625
Industrial Aptitude Spatial Orientation	7	456	361	471	453	724	661		341	704	522
Same-Opposite (ACE)	8	225	207	314	272	424	362	341		743	624
Completion (ACE)	9	306	350	444	352	580	506	704	743		709
Verbal Analogies (ACE)	10	432	483	476	524	496	625	522	624	709	

An *r* of .212 is significant at the five per cent level of confidence.
An *r* of .277 is significant at the one per cent level of confidence.

The problem-solving protocols were analyzed using the following procedure: (*a*) The investigator analyzed the protocols by reading the typed copies and listening to the recordings in an attempt to arrive at meaningful categories of problem-solving techniques. (*b*) Each hypothesis was checked by using the investigator's analysis of the protocols plus the answers which subjects gave during the third portion of the interview. (*c*) The investigator's classification of protocols as to categories was checked by two independent judges, one psychologically sophisticated, the other psychologically naive. In all instances agreement between judges and investigator was significant beyond the one per cent level of confidence. (*d*) A verb count was made of the protocols relative to the operational definitions of the various factors.

Using the above type of analysis of the protocols, the investigator and judges agreed that the hypothesized definitions of the space factors were accurate, yet not adequate in explaining the processes involved in solving the problems. Thus, after studying the problem-solving protocols and on the basis of the verb counts, the following slightly revised operational definitions were formulated.

TABLE 2

MATRIX CONTAINING CENTROID FACTOR LOADINGS

Test		Factors				h^2	
		I_c 1	II_c 2	III_c 3	IV_c 4	Obtained 5	Original estimated 6
Flags	1	719	496	—153	—175	817	839
Figures	2	704	510	—295	—154	866	839
Cards	3	723	262	—054	—092	603	653
DAT Space Relations	4	662	143	253	208	566	591
Chair-Window	5	757	—134	291	—176	707	724
Guilford-Zimmerman Spatial Orientation	6	765	066	214	168	664	661
Industrial Aptitude Spatial Orientation	7	752	—194	326	—279	787	724
Same-Opposite (ACE)	8	591	—458	—342	192	713	743
Completion (ACE)	9	755	—537	—207	—153	925	743
Verbal Analogies (ACE)	10	777	—179	—136	270	727	709

Decimal points omitted. All entries are to three decimal places.

The *first space factor* represents the mental ability to *turn* or *rotate* a given figure or part of that figure in one plane (or about an imaginary axis) to see if it corresponds to another figure in the same plane. (Eighty-

TABLE 3

MATRIX CONTAINING ROTATED FACTORS

Test		Rotated Factors				h^2	
		I_2	II_3	III_3	IV_2	Rotated	Centroid
		1	2	3	4	5	6
Flags	1	000	807	099	400	821	817
Figures	2	034	874	000	325	871	866
Cards	3	150	615	174	413	602	603
DAT Space Relations	4	247	240	145	653	566	566
Chair-Window	5	350	295	556	435	707	707
Guilford-Zimmerman Spatial Orientation	6	355	309	210	636	670	664
Industrial Aptitude Spatial Orientation	7	347	280	671	383	796	787
Same-Opposite (ACE)	8	812	260	016	025	728	713
Completion (ACE)	9	781	370	420	000	923	925
Verbal Analogies (ACE)	10	677	355	035	392	739	727

Decimal points omitted. All entries are to three decimal places.

The slight discrepancy between the communalities contained in Table 2 and Table 3 can be accounted for by the fact that the centroid axes were plotted on graph paper on which third place accuracy was a matter of the investigator's estimating distances between lines; however, this discrepancy was not large enough to indicate any appreciable error in the rotational procedure.

two of the 84 subjects used a method that fits this definition; the two who did not tried to use angles and figural cues only, without rotating the problems.)

Following are two excerpts from protocols illustrating the use of the method set forth in the above definition (responses to experimental items from Figures Test):

> *S* 16: ". . . I would look at these various choices here, and I would take the problem and try to switch it around, turn it around in the same form as these here; after I turn it around, I see that they can be made to coincide. . . ."
>
> *S* 44: "I'm trying to turn the figure around in a way that it is in the same position that the key problem would be."

The *second space factor* represents the mental ability to *see* or *observe* the spatial relationship of objects involved in dynamic situations, spatial relationships in which the subject has to imagine that the object or objects involved changed their positions in space relative to one another. (Depending on the difficulty level of the items being considered, somewhere between 76 and 83 subjects used such a method.)

Examples (responses to experimental items from the *DAT* Space Test):

> *S* 11: "Well, I folded it together, and obviously two of the ends were black, so A is wrong, etc. . . . You fold the thing and it's a box."
>
> *S* 6: ". . . well, if this pattern were folded into an object, it would be a square with two opposite sides colored. . . ."

The *third space factor* involves the ability to determine "from where you are looking at an object"; i.e., where one is spatially located in relationship to a particular object. (On the Industrial Aptitude Spatial Orientation Test, 58 of the subjects used a method similar to this definition.)

Example (response to experimental items from the Industrial Aptitude Spatial Orientation Test):

> *S* 2: ". . . I'm just presuming that I'm walking around the table, a tennis table, and the view I'd get of it as I'm walking around that table. . . ."

In addition to the formulation of these definitions, the following observations were made:

1. In analyzing the problem-solving protocols for the Figures Test, which had the highest loading on the first space factor, it was noted that the "part vs. whole" approach was another way the protocols could be classified; of the 82 subjects who used the approach stated in the revised definition, 39 rotated the whole figure, 43 rotated or turned only part of the figure. Subjects who made the highest scores on this test used the "part" approach more often; this difference was significant beyond the one per cent level of confidence.

An example of using the "part" approach:

> *S* 63: "What I did was orient myself by some identifying mark as the curved line at the bottom; then I noticed another marking symbol as either going to the left or right when I had it in this position. . . ."

An example of using the "whole" approach:

> *S* 18: ". . . I'd work it out the same way. I'd try to put the figures on top of one another and twist them around. . . ."

2. On the Figures Test, subjects who made higher total scores did not try to relate the figures to something familiar (or concrete), but rather used abstract symbolism. This difference was also significant beyond the one per cent level of confidence.

Examples of using familiar reference objects:

> *S* 37: "Well, first of all looking at the practice problem, I would notice there is sort of a J, a backward J, and a vertical V."
>
> *S* 46: "That one looks like the right half of a keyhole, and that's the system I used."

An example of using abstract rather than familiar objects:

> *S* 68: "First you switch the figure around mentally, take a mental picture of the complete figure. . . ."

3. In analyzing the *DAT* protocols (the *DAT* Space Test had the highest loadings on the second space factor), it was noted that there were four sub-categories of problem-solving processes used and that these processes were used with significantly different frequencies depending on the difficulty level of the items. The four sub-categories of problem-solving processes were:

(*a*) Subject spontaneously folded pattern and then noted relationship of parts (on the easier problems 57 subjects used this method; on the more difficult problems, only 12). An example of using this method is as follows: *S* 67, ". . . first of all the block is folded up; try to picture the block as it would be. . . ."

(*b*) Subject looked first at response figures (solid objects to the right) and then looked at the pattern, may even have reported unfolding one of the response figures to see if it was the same as the pattern (on easy problems 17 subjects used this method, on more difficult problems, 20). Example: *S* 18, "First of all I would look at the shape of the material, look at the shape of the finished product over here, and see if they could possibly look anything alike. . . . I looked at the finished product to set up in my mind how these problems or patterns could be folded into finished products."

(*c*) Subject did not fold or unfold stimulus pattern or response figures at all but looked entirely for other cues such as angle intersection, etc. (on easy problems 7 subjects used this method, difficult 44). Example: *S* 16, "By following the right-hand corner, it has a long part here, as long as it is square right here; *A* is correct on account of having this long seat here. . . ."

(*d*) Subject who guessed, could not work the problems (on easy problems, one subject did this; on difficult problems, eight subjects). Example: *S* 27, "Well, I was confused on that (*DAT* Space Test); I never did know if the shaded part had to be next to the shaded part shown, and I don't think I got any of those correct. . . . Well, I was confused on this test; I just more or less had to guess."

X^2 for difference in the use of methods on the easy vs. difficult items was 61.75. Davis' method (2) was used for determining difficulty level of items; difficulty level of easier item was 78; difficulty level of more difficult item was 36.

4. In analyzing the protocols for the Guilford-Zimmerman Spatial Orientation and the Industrial Aptitude Spatial Orientation Tests, it was noted that two sub-groups of problem-solving methods were being used:

(*a*) Subjects imagined themselves being reoriented with regard to the problems; i.e., they imagined themselves looking at a certain figure or object from different angles—they were an inherent part of the problem (26 subjects used this approach on the Guilford-Zimmerman Spatial Orientation Test, 58 on the Industrial Aptitude Spatial Orientation Test). Examples: *S* 2 (Guilford-Zimmerman), ". . . if I imagined myself standing in a rowboat, and if you push down with your left foot, and then turn your boat this way, it gives the appearance of the horizon at this point out here. . ."; *S* 27 (Industrial Aptitude), ". . . well, you'd have to be standing right here in order to see the little ball being the shortest distance from the center on your right-hand side."

(*b*) Subjects rotated and moved stimulus and response problems but did not imagine themselves being reoriented—they were not an inherent part of the problems (58 subjects used this approach on the Guilford-Zimmerman Spatial Orientation Test, 26 on the Industrial Aptitude Spatial Orientation Test). Examples: *S* 8 (Guilford-Zimmerman), "I do it by setting up two different planes; the boat is always constant in the picture, and it's the background that changes. On this one I held the background steady and brought the boat up at another angle . . ."; *S* 63 (Industrial Aptitude), "First of all I notice the three pool balls; two of them are lined up closer than the other; that large stripe goes through the middle one and perpendicular to the line of the three; then I notice that Figure 13 is all right because the large band is perpendicular to the line of these three balls."

X^2 for the difference in use of the two methods on the Industrial Aptitude Spatial Orientation and Guilford-Zimmerman Spatial Orientation Test was 32 (significant beyond the one per cent level of confidence). It was noted earlier that the Guilford-Zimmerman Spatial Orientation Test was the only test which did not appear on the space factor (Space 3) that it was predicted to measure in this investigation. This analysis of problem-solving protocols offers a possible cue as to why it did not—the majority of subjects were solving the problems the same way they solved those on the *DAT* Space Test, which was typical of the second space factor. In the factor analysis, the *DAT* and Guilford-Zimmerman Spatial Orientation Tests both had high factor loadings on the second space factor, while they had low loadings on the third space factor. Although this could possibly have been the result of not having enough tests to represent the second space factor, it seems reasonable that it could also have been a result of the subjects' use of different problem-solving methods.

E. SUMMARY

In the present study, operational definitions of three space factors were formulated on the basis of the analysis of the subjects' problem-solving techniques. Also, some specific techniques used in solving problems on tests typical of the various factors were observed.

Further investigations should be conducted: (*a*) to decide how important these various techniques are in determining one's score on specific tests, (*b*) to determine if test scores can be raised appreciably by teaching subjects different problem-solving techniques, and (*c*) to determine what problem-solving techniques subjects are using to solve tests in other ability areas.

REFERENCES

1. BLOOM, B. S., & BRODER, L. J. Problem Solving Processes of College Students. Chicago: Univ. Chicago Press, 1950.
2. DAVIS, F. B. Item-Analysis Data. Cambridge, Mass.: Graduate School of Education, Harvard University, 1949.
3. DEWEY, J. How We Think. New York: Heath, 1935.
4. GUILFORD, J. P. (*Ed.*). Printed Classification Tests. Army Air Forces Aviation Psychology Program, Report 5. Washington: U. S. Government Printing Office, 1947.
5. KELLEY, T. L. Crossroads in the Mind of Man. Stanford: Stanford Univ. Press, 1928.
6. THURSTONE, L. L. Some Primary Abilities in Visual Thinking. *Psychometric Laboratory Report* 59. Chicago: Univ. Chicago, 1950.
7. TITCHENER, E. B. Lectures on the Experimental Psychology of the Thought Process. New York: Macmillan, 1909.
8. WERTHEIMER, M. Productive Thinking. New York: Harper, 1945.
9. WOLFLE, D. Factor Analysis to 1940. *Psychomet. Monog.*, 1940, No. 3.

Department of Psychology
University of Delaware
Newark, Delaware

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 27-49.

THEORY AND TREATMENT OF STUTTERING AS AN APPROACH-AVOIDANCE CONFLICT*[1]

Department of Psychology, University of California

JOSEPH G. SHEEHAN[2]

A. THE PROBLEM

Seldom is the rôle of conflict, so important in every problem of adjustment, so clearly portrayed as in the disorder of stuttering. Stuttering behavior is essentially a hesitancy, an interruption in the forward flow of speech, a holding back in a situation which calls for going ahead. In the author's view, this hesitant or avoidant aspect of stuttering reflects the fundamental nature of the disorder:

> Stuttering is a result of approach-avoidance conflict, of opposed urges to speak and to hold back from speaking. The "holding back" may be due either to learned avoidances or to unconscious motives; the approach-avoidance formulation fits both (45).

This is, briefly stated, the theory to be presented in this paper. It seeks to integrate advances in speech pathology, psychopathology, and learning theory into a systematic theory of stuttering.

In recent years Miller (40) and Dollard and Miller (10) have formulated in systematic terms the nature of conflicts, especially those involving simultaneous approach and avoidance tendencies. Lewin (36) analyzed conflicts between driving and restraining forces in relation to positive and negative valences. Travis, (53, 54) Glauber (24), and Fenichel (16) have developed psychoanalytic theories of stuttering based in part on conflict between conscious and unconscious wishes. Johnson (17, 30), Bryngelson (8), Van Riper (60, 61), and speech pathologists of the Iowa tradition have evolved, under different theoretical banners, systems of treatment which stressed non-avoidance and reduction of fears.

Aspects of conflict in stuttering have been noted by a number of investigators. More than 80 years ago the German writer Wyneken called the stutterer a

---

*Accepted for publication by Roy M. Dorcus of the Editorial Board, received in the Editorial Office on January 22, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Based on papers delivered at the meetings of the Western Psychological Association, April, 1950, and the American Speech and Hearing Association, November, 1950.

[2] The author wishes to express his indebtedness to his teachers, Drs. C. Van Riper and Wendell Johnson, whose influence is apparent beyond their many acknowledgments throughout the paper.

Sprachzweifler, a "speech-doubter," comparing his plight to that of one seized with uncertainty at the moment of attempting a leap (63). In 1936 Johnson and Knott (32) defined stuttering as the manifestation of a conflict between the communicative drive or impulse and the impulse to inhibit expected stuttering. Knott and Johnson noted ". . . an experimental conflict . . . marked by the attempt of the stutterer to perform two mutually opposed tasks simultaneously; one is to talk, the other not to stutter (34).

As part of his analytically oriented theory, Travis held stuttering to be "a compromise between 'letting out' and 'holding in' (54). . . The stutterer wishes to express himself, and at the same time fears doing so" (53). Fenichel stated: "The symptom of stuttering reveals a conflict between antagonistic tendencies; the patient shows us that he wishes to say something and yet does not wish to" (16, p. 311). Closely related observations of conflict in stuttering may be found in Blanton and Blanton (3), Bluemel (5, 6), Dunlap (11), Fletcher (18, pp. 231-232), Froeschels, (22, 23), Glauber (24), Hill (26), Johnson (29, 30, 32), Solomon (51, 52), Van Riper (57, 60, pp. 269, 277, 287, 321-322, 61, pp. 19-20), Wischner (62, p. 330), and Wyneken (63).

Wide divergence has appeared among various writers as to the nature of the conflict and its relation to the stuttering. For some it has been a conflict over gratification of instincts, for others a conscious interference with an automatic process, for still others a rivalry between cortical hemispheres. Many apparently competing theories attack the problem at different levels and do not necessarily contradict one another. In like manner, the approach-avoidance conflict theory presented here involves only its own level of analysis and is sufficiently broad to be compatible with many other interpretations of stuttering.

B. Two Essential Questions

If we reduce stuttering behavior to the simplest possible terms, we find that it is a *momentary blocking*. Almost mysteriously the stutterer is stuck on a word, and then, for reasons just as baffling, he is able to continue. An explanation of stuttering must account for these twin features of the stutterer's behavior.

Most theories of stuttering have focused on the hesitancy, on what produces the blocking. But from the standpoint of systematic theory as well as therapy, it is just as important to explain termination of the block as the block itself. This problem has been considered in an earlier study: "What seems to determine the moment of release, the moment at which the stutterer can finally say the word?" (44, p. 69).

Two questions then become essential in the explanation of the stutterer's behavior: (1) What makes him stop? (2) What enables him to continue? To answer them fully, with support for the answers, requires the complete spelling out of the theory portion of this paper.

The two central hypotheses, however, may be stated quite briefly:

(1) *The Conflict Hypothesis.* The stutterer stops whenever conflicting approach and avoidance tendencies reach an equilibrium.

(2) *The Fear-Reduction Hypothesis.* The occurrence of stuttering reduces the fear which elicited it, so that *during* the block there is sufficient reduction in fear-motivated avoidance to resolve the conflict, permitting release of the blocked word.

The discussion that follows will take up each of these in turn.

C. THE CONFLICT HYPOTHESIS

1. *Statement*

If stuttering occurs whenever approach and avoidance tendencies reach an equilibrium, we should be able to analyze the process in terms of relative strengths of gradients of each. Miller (40, 41) and Dollard and Miller (10) have provided us with an excellent theoretical model for such an analysis (Figure 1).

For the stutterer, the speaking of a difficult word involves a goal, that of communication, but also a fear, that of inability to communicate. The stutterer thus has a "feared goal" in Miller's sense. Conflicting tendencies in the stutterer to approach and to avoid are represented by solid and broken lines

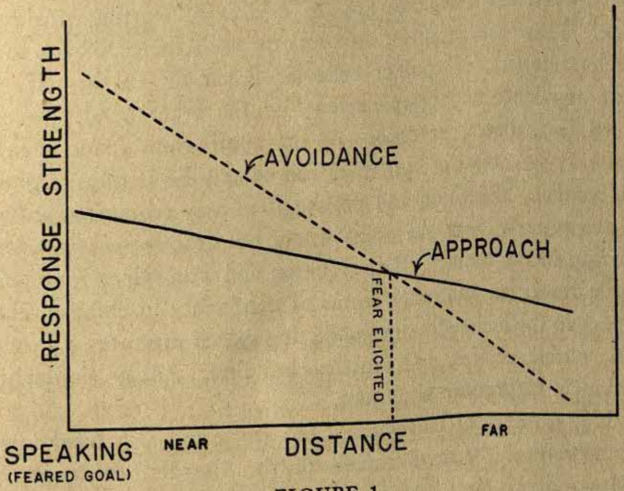


FIGURE 1

respectively. From the fact that the fear-motivated avoidance gradient is steeper than the reward-motivated approach gradient, it can be seen that an organism put in an approach-avoidance conflict situation will *go part-way and then stop,* or oscillate helplessly in the zone where the gradients cross. This is exactly the behavior the stutterer shows in attempting a feared word, or upon entering a feared situation. He says "K-K-K-Katy" or blocks silently after having begun the word. He freezes at the instant of picking up the phone, or halts on the threshold of a strange office.

If this formulation is essentially correct, we are in a position to answer the first question. The stutterer stops after advancing part-way because he is in a conflict situation, and the moment of his stopping is determined by the relative strengths of approach and avoidance gradients. Stuttering behavior itself has a hesitant character because it is the result of a conflict. Such an interpretation of stuttering accords well with Freud's classic view of the nature of neurotic conflict:

> ". . . neurotic symptoms . . . are the result of a conflict. The two powers which have entered into opposition meet together again in the symptom and become reconciled by means of the *compromise* contained in the symptom-formation" (19, p. 313).

In the compromise, i.e., the symptom of stuttering, the conflict is neatly externalized.

Many of the secondary symptoms of stuttering, as analyzed by Van Riper (57), may be interpreted as compensatory efforts to overcome avoidance, to go forward in the face of fear. Others, as he has pointed out, are just reactions to fear, directly expressed avoidance. Among the compensatory measures to overcome avoidance are the devices of starting, antiexpectancy, and release. These are, in effect, attempts to reach the goal by a roundabout route, a characteristic of conflict behavior noted by Lewin (37, p. 263). Devices of avoidance and postponement, on the other hand, are essentially reactions to fear, direct expressions of the approach-avoidance conflict.

The primary symptoms of stuttering can equally be attributed to approach-avoidance conflict. Repetition and prolongation may represent oscillating and stopping, respectively, near the point where the gradients cross, at the point of equilibrium in the conflict. Both Johnson and Van Riper have pointed out that syllable repetition and other forms of nonfluency in children characteristically appear in the midst of many conflicting speech pressures, e.g., vocabulary acquisition, phrase choices, and sentence building. Adults similarly tend to speak hesitantly in pressure situations.

Speech is a sequence of movements, and stuttering is a breakdown in the sequence. Whatever the involvements of the disorder, the point at which the breakdown occurs must bear a relation to these involvements. The breakdown occurs early in the sequence, but seldom prevents initiation of the sequence. Stuttering is, in other words, chiefly a disorder of *release,* of going partway and then stopping. Froeschels (22) has called repetition and prolongation ("clonus and tonus") the only symptoms common to all stutterers, an observation supported by the author's findings (44, p. 62). The stopping and oscillation which these symptoms may express are the features common to approach-avoidance behavior.

For clarity in presentation, this discussion has centered on stuttering as simple approach-avoidance conflict. Speaking has been the approach response; not speaking the avoidance response. But at times there are approach and avoidance tendencies to the act of speaking itself as well as to the act of not speaking. This involves some additional assumptions[3] and a further analysis, in terms of Miller's double approach avoidance conflict (Figures 2 and 3), which may account better for certain occurrences of stuttering.

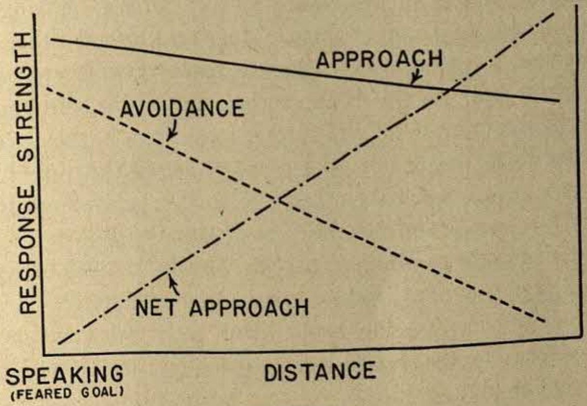


FIGURE 2

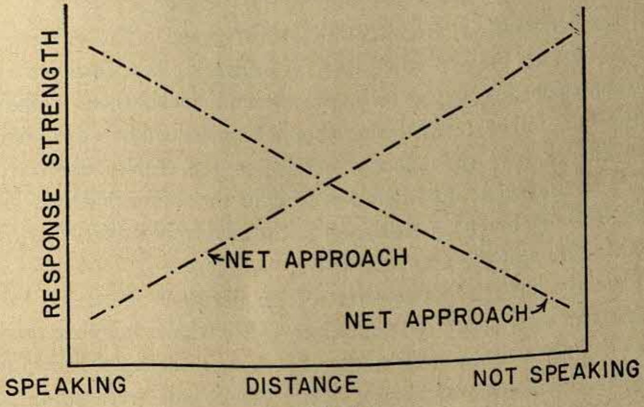


FIGURE 3

---

[3] The writer is indebted to Dr. Irving Maltzman for portions of this analysis.

The conflict on speaking is as follows: There is an approach tendency for speaking, since it is socially demanded. But since speaking entails the danger of stuttering, there is an avoidance tendency based on the fear elicited by this danger. This is the type of conflict which Johnson and Knott (32) held responsible for stuttering.

The conflict on not speaking is as follows: There is an approach tendency for not speaking, because silence becomes an attractive alternative to the danger situation of speaking. But this alternative also is to be feared.

In a situation which calls for speech, not speaking or not being able to speak in itself involves a threat. Many stutterers show a *fear of silence,* and any dead stop in their communication spurs panicky efforts to release the block. Many of the irrelevant and apparently unintelligent symptoms of the stutterer can be understood as a filibustering, a measure taken against the fear of silence.

Consequently, there is an avoidance tendency for not speaking as well as an approach tendency for not speaking. Movement toward either feared goal elicits more fear, so that the net approach tendency will be greater toward the more distant goal. As this is approached, e.g., as the stutterer gets closer to the speech attempt or as he gets further away from it, the alternative goal becomes more distant, hence less dangerous and more attractive. The stutterer will then turn and approach the other goal until it becomes too feared. Dollard and Miller's prediction for this type of situation is "one of stable equilibrium like that of the pendulum" (10, pp. 366-367). Such behavior toward feared words and situations is characteristic of the stutterer.

The paradoxical increase in "fear elicited" as lowered avoidance permits a closer approach to the feared goal is equally true of simple approach-avoidance conflict (1).

2. *Evidence*

What predictions flow from conflict theory, and what evidence can be brought to bear on these predictions? If stuttering is a form of conflict, a resultant of competing urges to approach and avoid, then it should vary systematically as follows: Stuttering should be increased by (*a*) a heightening of the avoidance drive through an increase in the penalty upon which fear and avoidance are based; (*b*) by a lowering of the approach drive. Stuttering should be decreased by (*a*) a reduction in the avoidance drive (fear, penalty); (*b*) by an increase in the approach drive.

What are the findings? The effect of penalty upon stuttering was shown in an important early study by Van Riper (58), who found that frequency of stuttering is increased by expectation of electric shock to be administered according to the number of blocks. Trial shocks only were administered, so that Van Riper's procedure was essentially a shock threat situation, a neat experimental paradigm for the relationship between stuttering and fear. Recently these results have been repeated by Frick as part of a comprehensive investigation of the relationship between punishment and stuttering. In other portions of his experiment, Frick administered shock for each stuttered word, in one condition at the end of the passage and in another immediately after the block. In each case the added penalty resulted in greater frequency of stuttering (21). Porter found that the frequency of stuttering is increased by greater social penalty, as measured by the number of persons in the audience (43). On the other hand, the frequency of stuttering is decreased by a reduction in penalty as measured by the meaningfulness of the material (Eisenson and Horowitz, 13), and by degree of communicative responsibility (Eisenson and Wells, 14). Bloodstein (4, p. 296) has suggested that the adaptation effect, the reduction in frequency of stuttering with repeated readings of the same passage, may be explained on the basis of a reduction in propositionality. As in the Eisenson studies, this involves a reduction in penalty. These findings accord well with the predictions of the approach-avoidance theory.

So much for penalty, and increases or decreases in the avoidance drive. What about the approach drive? The evidence on this is sparse. That which we do have, mostly clinical observation, supports our formulation. Stutterers, like people generally, show markedly more hesitancy when dealing with material they are reluctant to reveal. Literature is full of descriptions of small boys stammering their excuses and of the hesitant speech of marriage proposals. Stuttering has been known to start from a parental demand of an oral confession of guilt. The reduction of communicatory drive in such a situation is fairly obvious. On the other hand, increases in the approach drive brought about through hypnosis, suggestion (4) or speaking under strong or unusual stimulation (4) brings increased fluency. Stutterers sometimes surprise themselves and their associates by speaking normally in a crisis situation. Anger or sudden stress may produce a spectacular fluency. Such events point to the effect of a temporary increase in the approach drive.

A recently completed study bears directly on the strengthening of approach behavior in stutterers. Twenty adult stutterers read two 200-word passages five consecutive times under different conditions. The control condition was a standard adaptation situation, in which the subject simply read the passage each time as naturally as possible. Under the experimental condition, the subject repeated each stuttered word until he had attacked it once successfully. This technique led to more normal speech and less stuttering ($P = .02$). Under the assumptions stated in the study, the approach response of speaking was thereby strengthened while the avoidant response of stuttering was moved further away from the point of reinforcement and correspondingly weakened (46).

3. *Levels of Conflict*

In order to present the conflict hypothesis directly and simply, more thorough analysis of the different levels at which conflict may occur in stuttering has been reserved for separate presentation. Five distinct levels emerge: word-level; situation-level; emotional content level; relationship level, and ego-protective level.

At the *word level* the conflict is between the urge to speak the word and the urge not to speak the word. For example, the stutterer wants to say "hello," i.e., he approaches this word, but fear holds him back, because "h" sounds have through past experience acquired cue value.

At the *situation level* there is a parallel conflict between entering and not entering a feared situation. The stutterer's behavior toward using the telephone, reciting in classes, or introducing himself to strangers illustrates this conflict. Many situations which demand speech hold enough threat to produce a competing drive to hold back.

These two levels are based upon Van Riper's distinction between specific expectancy, or word-fear, as contrasted with general expectancy or situation-fear (57, 60). Such fears are usually based on past speech experiences and involve directly learned avoidance. How conflict based on word fear and situation fear can lead to blocking has been illustrated briefly by our diagram, Figure 1, and analyzed in detail elsewhere (46).

Conflict due to the *emotional content* of words, apart from their phonetic properties, such as has been described by Travis (53, p. 202), involved unconscious motivation for avoidance. This is illustrated by the stutterer whose speech grows worse in describing a traumatic experience, or upon giving information he is reluctant to divulge. It is the emotional and verbal content that is involved, not the situation or the words in themselves. Dunlap (11) said of the stutterer: "Sometimes he is in constant fear lest he inadvertently reveal something he would rather his elders did not hear." The inhibition of emotional content, especially hostile feeling, is a common denominator in many instances of stuttering (1).

At a closely related level of conflict, the occurrence of stuttering is in part a function of the *relationship* between the stutterer and his listener. Some stutterers experience no fear when they play a dominant rôle. One may give a fluent public address but stutter to individuals in the audience before and afterward. A child may block severely before one parent but speak easily to the other. An army enlisted man could never say "sergeant" until he became one. The fact that many stutterers can act in plays seems to show the effect of changed rôle. Several of the conditions listed by Bloodstein (4) as involving the behavior of the listener may equally be interpreted in terms of relationship. Pointing out that the stutterer can usually talk well when alone, Adler concludes, ". . . I can only interpret his stammer as the expression of his attitude toward others" (2, p. 63). Fletcher (18), Travis (53), Fenichel (16, pp. 311-314), Abbott (1), and many others have stated in one way or another the importance of the relationship between the stutterer and his listener.

Since stuttering may be termed a compromise between speech and silence, we should consider the symbolic significance of all three in analyzing what is expressed toward the listener. Depending upon our initial reference points, we may get widely differing results (*a*) stuttering may be considered an act of aggression against the listener; fluency then would be nonaggressive or friendly behavior; (*b*) speech may be considered an aggressive, competitive, phallic (in psychoanalytic terms) act; silence here is presumably neutral; (*c*) silence in itself may be viewed as hostile. Noting that in dreams to be mute is a symbol of death, Fenichel (16, pp. 311-313) likened stuttering to a partial mutism resulting from a turning inward against the stutterer's own ego of hostile urges originally directed against the listener. From his discussion it is possible to infer all three of the above meanings. Perhaps it should be concluded simply that either stuttering or speech or silence can be used to express hostility or more positive feelings at various times, but that no constant or universal meaning can be offered. Whatever the behavior through which the feeling is expressed, the occurrence of stuttering will in part be determined by degree of conflict at the relationship level.

At the *ego-protective level*, stuttering serves as a lifelong defense mechanism keeping its possessor out of dangerous competition. Through the stuttering, certain aspirations may be abandoned which might involve threat of failure, or threat of success.

Although our formulation of the ego-protective level stresses life aspirations and goals, the concept is sufficiently broad to include any defensive function of stuttering. Travis' excellent formulation more nearly represents the classic Freudian view of symptoms as defenses against forbidden impulses:

> ". . . stuttering is a defense created with extraordinary skill and designed to prevent anxiety from developing when certain impulses of which the stutterer dares not become aware, threaten to expose themselves" (53, p. 193).

That stutterers do tend significantly to avoid threat of failure is shown by their behavior in an experimental level of aspiration situation. Compared to normal controls, stutterers set less exacting (though more realistic) goals for themselves, predicted more modest performances, and showed in general a lower level of aspiration (49).

Any particular moment of stuttering may be understood in terms of the interplay of conflicting forces at these levels. Whatever the level of conflict, the overt conflict is experienced at the word-level. Conflict at the ego-protective level, for example, might explain a stutterer's resistance in therapy. It would not explain his blocking on a word, except through the mediation of word-fear. In like manner, the pressure of the situation, emotionality of the utterance, or nature of the interpersonal relation must ultimately be expressed at the word-level. It is in this fundamental sense that stuttering is a conflict between speaking and not speaking.

D. The Fear Reduction Hypothesis

1. *Statement*

If stuttering results from a conflict, how is the conflict resolved? According to the conflict hypothesis, the stutterer stops when approach and avoidance tendencies reach an equilibrium. Why then doesn't the stutterer remain permanently in a conflict situation? How is he able to obtain release from the block?

At the beginning of a block the stutterer is stuck. He cannot get the word out. By the end of the block some change has taken place, so that he can now utter the word. What has happened between these points? If the stutterer cannot say the word for a time, why is he able to say it at all? To account for this, the hypothesis is advanced that the occurrence of stuttering brings about a reduction of the fear which elicited it.

During the moment of stuttering, there must be sufficient reduction of fear, avoidance tendency and conflict to "release" the blocked word. Were it not for this fact, once the stutterer became stuck on a word, he would remain stuck indefinitely.

How can the occurrence of stuttering reduce the fear that elicits it? Why should the stutterer be in any better position to speak the word at the end of a block than at the beginning? After all, the symptoms of stuttering frequently bear little relevance to the actual speaking of the word (57, 44).

The occurrence of stuttering may effect a reduction of fear in several ways:

First, since the stutterer's fear is tied up with his avoidance and is an effort to hide the disapproved symptoms, once the stuttering begins to occur it can no longer be hidden. The stuttering block itself forces the stutterer to "face" his stuttering to a certain extent. This reduces the conflict. The thing the stutterer feared has now occurred. He can no longer avoid it completely, and this partially reduces the fear. On the other hand, successful avoidance builds up tension.

Second, once the stuttering block begins to occur it is a known entity. As the stutterer begins to approach a block, he has a vague dread, a generalized expectancy. As he nears the actual moment of stuttering, he recognizes more familiar landmarks (initial sound, length of word, cf. Brown, 7) and is even able to predict the duration of the block with some precision (study by Milisen and Van Riper, 39). Hence, he gets some feeling of control, the fear is now more specific and there is less fear of the unknown, and there is a reduction of the element of fear from a sense of helplessness (study by Mowrer and Viek, 42).

Third, to the extent that stuttering can be interpreted as an aggressive act directed against the listener (Fenichel, 16) the stuttering relieves the aggression (frustration-aggression hypothesis), hence reduces the inhibition to aggression, hence reduces the approach-avoidance conflict of the stuttering.

To test these corollary hypotheses is obviously beyond the scope of this paper. They are stated here for the completeness of the theory and as problems for future research.

2. *Evidence*

Among studies which may be interpreted as revealing the effect of stuttering on fear are the following:

1. The study by Meissner (38) in which voluntary non-fluency reduced stuttering.
2. The studies of Johnson and Knott (33), Johnson and Inness (31), Shulman (50), and other studies showing the adaptation effect. From the fear-reduction hypothesis, we may infer that the occurrence of stuttering in the first reading dissipates sufficient anxiety so that there is less stuttering in the second reading, etc. Such an interpretation of adaptation is in agreement with and in large part contained in Johnson's current published view:

> The adaptation effect would appear to be a particularly clean-cut example of the effect of improved adjustment to the reading situation, lowered anxiety about stuttering resulting from "doing the thing feared" and discovering on the spot that the dreaded consequences turn out to be less dreadful than expected (30, p. 210).

3. The study by Brown (7) showing that stutterers tend to stutter more on the first word in a sentence. According to our hypothesis, the stuttering which occurs on the first word will, through generalization, make subsequent words in that sentence easier.

4. The effectiveness of the technique of negative practice as reported by Dunlap (11, 12), Case (9), and Fishman (17). Duplicating the stuttered performance voluntarily should "satisfy" the fear, lower the avoidance drive, and diminish the conflict causing the stuttering.

5. Van Riper's study of thoracic breathing (56) in which stutterers show a rehearsal of the block during an expectancy period. The rehearsal behavior is here interpreted as an effort to reduce the fear by stuttering subvocally first.

6. The finding of Travis, Tuttle, and Cowan (55) that during tonic stuttering block, heart rate becomes progressively slower.

7. Van Riper's study showing the relation of tremors to perpetuation of and release from the stuttering block (59). Results showed a tendency toward decrease in the rate and amplitude of jaw tremors during the block.

8. In an effort to verify Van Riper's result, and as a specific test of the fear-reduction hypothesis presented here, the author and Mr. Robert Voas of U.C.L.A. carried out a study of physiological tension patterns of 15 stutterers during the moment of stuttering (47, 48). Masseter muscle action currents and other measures showed a rise in amplitude (tension) at the beginning of the block followed by a dropping off to the moment of release. Upon the speaking of the word—the reaching of the goal—the resolving of the conflict—a marked reduction in tension occurred as the muscle group returned to the resting state. It should be pointed out that this type of tension reduction occurs normally with the speaking of a word whether stuttered or not. The theoretically important feature is that sufficient fear reduction occurs *during* the stuttering block to permit release of the blocked word. Van Riper has analyzed some of the ways in which stuttering brings about fear-reduction:

> "The relationship of fear to stuttering is so great that stutterers often feel that, if they could destroy the fear of approaching words, they would be able to say them without any difficulty (60, p. 298). . . . Fear of a word brings an automatic reaction of the oft-practiced tricks previously used to reduce or counteract that fear" (60, p. 286).

The results of the study just described support this interpretation as well as Van Riper's unpublished findings in the tremor study.

9. Two other findings may be cited in support of the fear-reduction hypothesis. First, in applying a non-reinforcement procedure to a group of stutterers, it was possible to obtain a normal speaking of the word simply by asking the subject to repeat the word, i.e., to stutter on it until all fear present was "satisfied." That the subjects were able to do this, that they could speak the word more easily the second, third, and fourth attempt, is in itself strong evidence for a mechanism of fear reduction resulting from stuttering (46). In another study, (44, p. 25) a tendency was noted toward

stuttering with succeedingy larger portions of the word, e.g., "th-thir-thirty." From our hypothesis, the stuttering which occurred on the first attempt reduced fear and avoidance sufficiently so that succeeding points of stoppage occurred closer to the goal, i.e., nearer the speaking of the word.

3. *Implications*

These results have interesting ramifications. Van Riper states that successful avoidance increases the stutterer's fear and causes more trouble in the long run (60, pp. 284-285). The concept of stuttering as a fear-reducer clarifies this relationship. When the stutterer is able to avoid stuttering he does not dissipate the anxiety, the tension builds up and even though he may continue to be fluent, he is building up future trouble through the avoidance. On the other hand, when he stutters or pretends to stutter, he is reducing his fears and building toward greater fluency later on. Hence we have a paradoxical relation—the stuttering produces the fluency and the fluency produces the stuttering. This is a possible explanation of something frequently observed, namely, that stuttering tends to occur in waves.

We could thus view stuttering as having a function similar to that of tics, asthma, and many neurotic symptoms, that it "binds" the anxiety and has the property of reducing it. We would have a parallel here to compulsive acts. When the person cannot stutter or successfully does not, the fear builds up. Stuttering may be thus viewed as an expression of accumulated anxiety. When the behavior occurs, the anxiety erupts and is let out. This concept may go a long way toward explaining the apparently paradoxical fact that we can build up a stutterer's tension by asking him to be fluent and we can reduce his fear by encouraging him to engage freely in the expression of the symptom.

On the concept of anxiety binding Freud gives as one view, ". . . all symptom formation would be brought about solely to avoid anxiety. The symptoms bind the psychic energy which would otherwise be discharged as anxiety . . . if a compulsion neurotic is prevented from washing his hands after touching something, he becomes a prey to almost insupportable anxiety" (20, p. 85).

In the same context, the concepts of the stuttering symptom as a defense against anxiety has long been stressed by Travis (53, 54, p. 35).

It is as though each stutterer carries around with him a "stutter potential," a *reservoir of fear* which is tapped from time to time by the occurrence of stuttering. Such a concept, while more figurative than real, may help us understand the seemingly neurotic stutterer who suffers agonies in anticipation of stuttering that never quite happens. It is precisely because the individual never stutters that he never obtains relief. His need for stuttering, as Travis aptly put it, becomes acute. In a psychological sense he is much worse off than the individual who can express his conflict, and reduce his fear, through outward stuttering behavior. Such a concept has interesting possibilities: not only does it provide theoretical support for procedures used by Johnson and Van Riper with stutterers, but it suggests potentialities for *symptom expression* as a psychotherapeutic technique with any problem based on fear.

4. *Theoretical Relationships*

A paradoxical effect of fear-reduction during the block may be inferred from the nature of the conflict. By the lowering of the avoidance gradient, the stutterer is brought closer to the feared goal and hence may experience a paradoxical increase in "fear elicited." Thus a reduction in fear shifts the equilibrium of the forces in conflict and brings the stutterer closer to the speaking of the word, but simultaneously he is experiencing more fear as he nears the goal.

From the basic diagram of approach-avoidance conflict (our Figure 1), Dollard and Miller state their assumption that *fear elicited* is fixed by the height of the intersection of the two gradients, and proceed with the following analysis:

> . . . When the avoidance is strong relative to the approach, the two gradients intersect far from the goal. . . . At this distance the approach tendency is so weak that the subject is not strongly tempted to do frightening things. . . . Extinguishing the fear of the goal or making it seem less dangerous . . . will lower the entire gradient of avoidance . . . this will produce a paradoxical effect . . . as the apparent danger is reduced the subject will be more strongly tempted to do things which frighten him. Stronger fear and conflict will therefore be elicited (10).

At this point the situation must sound discouraging, but Dollard and Miller further point out:

> This deduction holds only for the range within which the gradients of approach and avoidance intersect. If the gradient of avoidance is weakened so much that it no longer intersects the gradient of approach, the subject will advance to the goal and further decreases in the gradient of avoidance will decrease the amount of fear and conflict (10).

This discussion may help us distinguish between the fear reduction *during* the block which permits release and that which results from the release, because the goal of speaking the word is reached. The fear reduction which determines the release may be obscured somewhat by a simultaneous increase in fear, conflict, and avoidance strength with the nearing of the feared goal.

Two shifts are occurring at once. For a given amount of fear and avoidance tendency, the point at which the stutterer initially stops is determined in part by the strength of the approach drive, which is in turn a function of his need to communicate and ability to tolerate anxiety. When he reaches the point of stoppage the stuttering begins to occur, reducing the fear which elicited it and lowering the avoidance gradient so that he moves closer to the goal. Thus the amount of fear subjectively experienced may remain relatively constant throughout the block. If the stutterer does not complete the block, but leaves the field or engages in some instrumental escape response, or if by using a device or trick he successfully reaches the goal in a roundabout manner, he does not "satisfy the fear" present and may continue to experience tension after the block. The unsatisfied fear may transfer over to a subsequent word and be dissipated on that word. When the opposite happens and there is rehearsal or sub-vocal satisfying of the fear first, there may be a stuttering on the word preceding the word originally feared, an observation made by Hill (26). Since fear-reduction reinforces the instrumental acts which bring it about, and reinforced responses tend to move up in the response sequence, (25, pp. 212-215) the stuttering tends to become anticipatory. Van Riper has shown that the actual form of the block may be rehearsed in breathing patterns prior to speech attempt. Through secondary reinforcement such rehearsal behavior will be fear-reducing, which indeed appears to be its function. Proprioceptive feedback from this rehearsal behavior enables the stutterer to anticipate his blocks, even to predicting their duration with some accuracy (39).

Thus there are indications of fear-reduction occurring during at least the first two of the three stages of the stuttering process as analyzed by Johnson: pre-spasm, spasm, and post-spasm (28). From fear-reduction occurring before and during the block and at the moment of release, the symptom is reinforced and maintained, the anxiety is "bound" within it, and a vicious circle is perpetuated.

E. TREATMENT

If stuttering is basically an approach-avoidance conflict, then the fundamental goal of treatment becomes the elimination of all tendency to avoidance, whatever the source. Since the avoidance is based on fear, stuttering is to be treated as a fear problem. Successful treatment then requires gaining a mastery over fear and finding expression of whatever needs, feelings, and tendencies have been locked up in the symptom. The approach-avoidance conflict theory of stuttering thus gives strong theoretical support to therapeutic procedures employed by Van Riper, Johnson, Bryngelson, Travis and others of the "Iowa School" of stuttering therapy.

Since approach-avoidance conflict theory was implicit in non-avoidance training and other techniques of the "Iowa School," the therapeutic approach presented here springs largely from this school. Two specific values of the approach-avoidance formulation should be mentioned: (*a*) Approach-avoidance theory relates fear, avoidance and conflict to stuttering, and theory to treatment in a systematic way. The paramount importance of reducing fears and avoidances becomes clearly apparent from the theory itself. (*b*) The analysis of levels of conflict and the distinction between learned avoidances and unconscious motives for avoidance provides for an integration of psychotherapy and speech therapy not always apparent in the writings of psychologists and speech pathologists.

The relative utilization of speech therapy and psychotherapy in present-day clinical work offers many contrasts. On one hand a few stragglers still use a purely symptomatic treatment which tries to ignore fears, prevent blocks, and build fluency directly through "confidence" measures. These people neglect all psychotherapy, and in a modern sense, speech therapy as well. At the opposite pole are many able clinicians, often psychoanalytically oriented, who insist that we must "ignore the symptom" and work toward the original cause, presumably something in the stutterer's past. For these people the speech pathologist could offer the stutterer nothing except in terms of whatever general psychotherapeutic abilities he might possess.

However, when stuttering is treated as a form of conflict, speech therapy and psychotherapy are not in competition, but have a common goal; the reduction of all tendencies to avoidance and of the fears which motivate them.

Although the stutterer's need for psychotherapy is now widely accepted, many of those who see this need most clearly continue to ask, "Why work on the speech at all?" To answer this question, and to defend the rôle that speech pathologists have come to assume in helping stutterers, we list the following points:

First, if we are really justified in viewing stuttering as an externalized conflict, should we not avail ourselves of this ready avenue to other levels? Conflict at the relationship level, for example, is mediated through situation and word-fear; why then not begin where the stuttering occurs and work back to the stutterer's relationships? The stutterer's blocks and his reactions to them mirror his self-concepts, attitudes toward others, mechanisms for handling conflicts, and many other aspects of his personality. Each time the stutterer has a block he is reliving in a small way one of the most significant experiences of his life.

Second, if we accept the hypotheses that stuttering reduces to some degree the fear that elicits it, possibilities suggest themselves for planned use of stuttering behavior as a tension reducer or outlet for fear. The "negative practice" techniques employed by Dunlap (12) probably owed some of their efficacy to this effect, though he advocated their use on different grounds. Such Iowa School methods as a voluntary repetition of "bounce," smooth prolongation or "slide," and duplication of the true pattern have been used successfully for years to reduce the stutterer's anxiety. Some of the best successes have been obtained with cases who had little outward stuttering to begin with and who underwent no discernible speech changes. But through such speech therapy, with very little conventional psychotherapy, they have been freed from the taxing strain of anticipation.

Third, profound personality changes can result from successful speech therapy. Teaching a stutterer how to handle his fears and his blocks, helping him develop healthier attitudes toward his problem, toward himself, and toward others—these things are not mere symptom treatment. Encouraging the stutterer to attack feared and difficult situations may support him strongly in something he really has wanted to do. Many stutterers relate that they get tired of running away; it is a deeply moving experience to master situations from which they have always retreated.

Fourth, a certain amount of direct speech therapy is necessary to deal with the maintaining causes of stuttering. Van Riper states, "Stuttering is peculiarly able to maintain itself once it gets started (60, p. 276) . . . in the secondary stage . . . the disorder has become 'chronic' or self-perpetuating . . . a vicious circle" (60, p. 239). Johnson observes: "Stuttering is something like our fear of thunder—there is no inherent or rational necessity for it, but once started it tends to be maintained, generation after generation" (30, p. 194). Both these authorities, as well as Dunlap (11, p. 199), Case (9), Wischner (62, pp. 330-331), and others, have pointed out how stuttering may perpetuate itself on a learned behavior, or habit level. Wischner considers stuttering as learned behavior which seems to resemble responses set up by instrumental avoidance training (62, p. 327), hence difficult to extinguish (62, p. 332).

An experiment by the author has shown that when reinforcement of the stuttering can be prevented, modified or decreased, stuttering is reduced. In this way the vicious circle may be broken (46, pp. 61-62). Van Riper's method of "cancellation," as well as a host of other techniques of the Iowa School, are directed toward the same result (60, pp. 348-371). With such techniques available, it is unnecessary to let the stutterer struggle helplessly with his blocks for an indefinite period while searching for buried complexes. And in long-range terms, if the therapist disdains all direct work on speech, he offers the stutterer only a purely verbal system for coping with long-reinforced habitual responses that are permitted to remain intact.

Treatment may begin at the topmost level of conflict and work downward. Conflicts due to immediate word fear and situation fear are dealt with through that specialized portion of the therapeutic process known as speech therapy. The speech therapy techniques appropriate to a conflict interpretation of stuttering have been so thoroughly developed in publications of Van Riper, Bryngelson, Johnson, and Travis as to require no repeating here. Conflicts at deeper levels are straight problems in psychotherapy, which may be handled concurrently. To the extent that such conflicts express themselves through outward stuttering behavior, they may be reached via the speech therapy. The stutterer's reaction to speech therapy frequently reveals the nature and extent of his resistances.

At the beginning the stutterer is caught in a vicious circle. He doesn't have the tools with which to work. Increasing his motivation at that point only increases the penalty on stuttering, therefore fear, conflict, and number of blocks. Speech therapy of the non-avoidance type can serve two important functions: first, for those whose stuttering has become self-maintaining or functionally autonomous, reducing drives toward avoidance may be sufficient to break up the "vicious circle" and free the stutterer; second, speech therapy gives the stutterer tools for dealing with word and situation fears he had not formerly possessed. When a stutterer rejects this opportunity he provides a focal point for the analysis and understanding of his own resistances. Such analysis is considered an essential first step in therapy by Fenichel (15, p. 46), who recommends that analysis of resistance precede analysis of content. This leads to the problem of dealing with unconscious sources of avoidance and conflict, i.e., psychotherapy for the stutterer.

Since the emotionality of the utterance is so frequently a source of conflict in the stutterer, an important goal in psychotherapy is *release* of *feeling*. Pent-up feelings or repressed emotional material responsible for producing blockings in speech require free expression and adequate outlet. Such goals are common in psychotherapy, and numerous techniques to achieve them are familiar terms in its literature: release of feeling, clarification of feeling, catharsis, abreaction, working through, play therapy, psychodrama, and others. It is interesting that one of the few modern therapists reporting fair success with stutterers is David Levy (35, pp. 77-79), whose "release therapy" is aimed primarily at the expression of feeling.

Many of the speech therapy techniques of the "Iowa School" may be interpreted as exercises in the expression of aggression. The stutterer who changes his mind about "faking" to a clerk because, "He seemed so nice, I couldn't do it to him," may illustrate a common underlying attitude. A "spluttering" or choking up of speech, behavior which at least superficially resembles stuttering, is a cultural caricature for inexpressible aggression. If stuttering may be viewed this way, in accord with Fenichel, then such techniques as negative practice and "bouncing" not only provide a means of abreaction but reduce the need to be aggressive. Such needs on the stutterer's part may be reduced in two ways: through the use of stuttering behavior itself, and through release of feelings behind it.

All maladjustments are in a sense due to a lack of satisfactory *relationships* in early life. People are neurotic to the extent that they lack good interpersonal relations. Psychotherapy, as the undoing process of maladjustment, must be directed toward establishing that which has been lacking—warm and satisfying relationships. In the beginning phases the therapist himself supplies the relationship. Later on, as treatment begins to be effective, the individual's improved adjustment is reflected in more congenial relations outside therapy.

Stuttering not only expresses the nature of the stutterer's relationships but is in part determined by them. When there is considerable conflict at this level, the working through of certain crucial relationships becomes essential to success. Fletcher showed awareness of this in regarding stuttering as a "social morbidity" (18, pp. 222-240), symptomatic of the victim's attitude toward all society. Recognition of the stutterer's need for better relationships has undoubtedly stimulated the development of group therapy methods with stutterers.

Dollard and Miller's analysis of the nature of approach-avoidance conflict may serve as a warning against bringing the stutterer too quickly into new and dangerous rôles. They point out that because of the paradoxical effect of a lowering of avoidance tendencies in eliciting greater fear, it is necessary to "dose anxiety" or to pace the individual so that he will not experience too much conflict at once and drop out of therapy (10, pp. 400-403). A further danger of reducing avoidance too quickly, and in all attempts at sudden cures, emerges from the ego-protective level of conflict.

At some point in the treatment of a stutterer it is necessary to consider the effect of the handicap, and possible recovery from it, upon goals and aspirations, upon what the individual has planned to do with his life. Specially constructed sentence completion test items, or open-ended questions, may elicit revealing projective material in this area. Examples: "What kind of a person would you be if you didn't stutter?" or "If your stuttering suddenly disappeared, what difference would it make in your life?" or "If I could get over stuttering, I would . . . ."

Frequently such questions make the stutterer anxious. Because the defeat may have become a peg upon which to hang all his shortcomings, or perhaps because of the capacity of the human organism to adapt itself to disturbance, the stutterer is likely to feel a little strange without his symptom. The individual may have lived with his stuttering so long that functioning without it involves a radical change in self-concept which must be gradual. To put it in another way, fluency may have become ego-alien. In the later stages of treatment the stutterer must learn to accept his new rôle with its fluency, just as in the early stages he needed to accept his old rôle with its stuttering.

As changes during treatment begin to occur, there should be a certain amount of *preparation for recovery*. Such preparation should include a careful psychotherapeutic exploration of the adaptiveness of the stutterer's goals, and the relation of these to the disorder. If the stutterer has enslaved himself to a striving toward unattainables, the therapist should help him find freedom. If the stutterer's level of aspiration is lower than his capabilities warrant, the therapist may help him realize the new possibilities opening up before him. All these have the effect of reducing avoidance and conflict due to ego-protective functions of the symptom.

From the foregoing, it may be seen that therapy is carried out at all five of the levels at which conflict occurs in stuttering—word and situation, feeling, relationship, and ego-protective. Psychotherapy and speech therapy are twin avenues to the common goal of reducing fear and avoidance. In the course of treatment, as well as in theory, the problem of stuttering may be handled and understood in terms of approach-avoidance conflict.

F. SUMMARY

1. Stuttering is a resultant of approach-avoidance conflict, of opposed urges to speak and to hold back from speaking.
2. The "holding back" may be due either to learned avoidances or to unconscious motives.
3. Principal hypotheses concerning stuttering behavior spring from two fundamental questions: (*a*) What produces blocking? (*b*) What determines release?
4. The Conflict Hypothesis: The stutterer blocks or stops whenever conflicting approach and avoidance tendencies reach an equilibrium.
5. The Fear-Reduction Hypothesis: The occurrence of stuttering reduces the fear which elicits it sufficiently to permit release of the blocked word, resolving the conflict momentarily and enabling the stutterer to continue.
6. In this way the symptom is reinforced and maintained, the anxiety is "bound" within it, and a vicious circle is perpetuated.
7. Conflict is directly expressed in primary symptoms of repetition and prolongation, which reflect a breakdown in the sequence of movements necessary to speech.

8. Secondary symptoms involve learned behavior representing compensatory efforts to overcome avoidance or to reach the goal by a roundabout route.

9. Conflict may occur at several levels—word, situation, emotional content, relationship, and ego-protective levels; any particular moment of stuttering is determined by interplay of forces at these levels.

10. Approach-avoidance theory relates fear, avoidance, and conflict to stuttering in a systematic way, so that goals of treatment become apparent from the theory itself.

11. Treatment proceeds through an integrated psychotherapy and speech therapy, aimed at attacking feared words and situations, releasing feelings, improving relationships, and freeing the individual from unadaptive goals, thereby achieving a total reduction of the fear and avoidance tendency responsible for the stutterer's conflict.

REFERENCES

1. Abbott, J. A. Repressed hostility as a factor in adult stuttering. *J. Speech Disor.*, 1947, **12**, 428-430.
2. Adler, A. Problems of Neurosis. London: Kegan Paul, Trench, Trubner, 1929.
3. Blanton, M. G., & Blanton, S. Speech Training for Children. New York: Century, 1924.
4. Bloodstein, O. N. Conditions under which stuttering is reduced or absent: A review of literature. *J. Speech Disor.*, 1949, **14**, 295-302.
5. Bluemel, C. S. Stammering and Allied Disorders. New York: Macmillan, 1935.
6. ———. Stammering and inhibition. *J. Speech Disor.*, 1940, **5**, 305-308.
7. Brown, S. F. The loci of stutterings in the speech sequence. *J. Speech Disor.*, 1945, **10**, 182-192.
8. Bryngelson, B., Chapman, M. E., & Hansen, O. K. Know Yourself: A Workbook for Those Who Stutter. Minneapolis: Burgess, 1950.
9. Case, H. M. Stuttering and speech blocking: A comparative study of maladjustment. Ph.D. dissertation, University of California at Los Angeles, 1940. Pp. 78.
10. Dollard, J., & Miller, N. E. Personality and Psychotherapy. New York: McGraw-Hill, 1950.
11. Dunlap, K. Habits: Their Making and Unmaking. New York: Liveright, 1932.
12. ———. Stammering: Its nature, etiology and therapy. *J. Comp. Psychol.*, 1944, **37**, 187-202.
13. Eisenson, J., & Horowitz, E. The influence of propositionality on stuttering. *J. Speech Disor.*, 1945, **10**, 193-197.
14. Eisenson, J., & Wells, C. A study of the influence of communicative responsibility in a choral speech situation for stutterers. *J. Speech Disor.*, 1942, **7**, 259-262.
15. Fenichel, O. Problems of Psychoanalytic Technique. Albany: Psychoanalytic Quarterly, 1939.
16. Fenichel, O. The Psychoanalytic Theory of Neurosis. New York: Norton, 1945.
17. Fishman, H. C. A study of the efficacy of negative practice as a corrective for stammering. *J. Speech Disor.*, 1937, **2**, 67-72.
18. Fletcher, J. M. The Problem of Stuttering. New York: Longmans, Green, 1928.
19. Freud, S. A General Introduction to Psychoanalysis. New York: Garden City Publ., 1943.

20. ———. The Problem of Anxiety. New York: Norton, 1936.
21. Frick, J. V. An exploratory study of the effect of punishment (electric shock) upon stuttering behavior. Ph.D. dissertation, State University of Iowa, Feb., 1951.
22. Froeschels, E. Stuttering and nystagmus. *Monatschr. f. Ohrenh.*, 1915, **49**, 161-167.
23. ———. Uber des Wesen des Stotterns. *Wien, Med. Wchnschr.*, 1914, **64**, 1067-1076.
24. Glauber, I. P. Psychoanalytic concepts of the stutterer. *Nerv. Child,* 1943, **2**, 172-180.
25. Hilgard, E. R., & Marquis, D. G. Conditioning and Learning. New York: Appleton-Century, 1940.
26. Hill, H. An interbehavioral analysis of several aspects of stuttering. *J. Gen. Psychol.,* 1945, **32**, 289-316.
27. Johnson, W. People in Quandaries. New York: Harper, 1946.
28. ———. An interpretation of stuttering. *Q. J. Speech,* 1933, **19**, 70-76.
29. Johnson, W. (and others) A study of the onset and development of stuttering. *J. Speech Disor.,* 1942, **7**, 251-257.
30. Johnson, W., Brown, S. F., Curtis, J. F., Edney, C. W., & Keaster, J. Speech Handicapped School Children. New York: Harper, 1948.
31. Johnson, W., & Inness, M. Studies in the psychology of stuttering: XIII. A statistical analysis of the adaptation and consistency effects in relation to stuttering. *J. Speech Disor.,* 1939, **4**, 79-86.
32. Johnson, W., & Knott, J. R. The moment of stuttering. *J. Genet. Psychol.,* 1936, **48**, 475-480.
33. ———. Studies in the psychology of stuttering: I. The distribution of moments of stuttering in successive readings of same material. *J. Speech Disor.,* 1937, **2**, 17-19.
34. Knott, J. R., & Johnson, W. An interpretive demonstration of ten observable facts about stuttering. In *Proceedings of the American Speech Correction Association.* Madison: College Typing Company, 1936, **6**, 150-154.
35. Levy, D. Release therapy. In Tomkins, S. S. (*Ed.*), *Contemporary Psychopathology.* Cambridge: Harvard Univ., 1947.
36. Lewin, K. Dynamic Theory of Personality. New York: McGraw-Hill, 1935.
37. ———. Field Theory in Social Science. New York: Harper, 1951.
38. Meissner, J. H. The relationship between voluntary non-fluency and stuttering. *J. Speech Disor.,* 1946, **11**, 13-23.
39. Milisen, R., & Van Riper, C. A study of the predicted duration of the stutterer's blocks as related to their actual duration. *J. Speech Disor.,* 1939, **4**, 339-345.
40. Miller, N. E. Experimental studies of conflict. In Hunt, J. McV. (*Ed.*), *Personality and the Behavior Disorders.* New York: Ronald Press, 1944. (Vol I, pp. 431-465.)
41. Miller, N. E. Comments on theoretical models illustrated by the development of a theory of conflict. *J. Personal.,* 1951, **20**, 82-100.
42. Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *J. Abn. & Soc. Psychol.,* 1948, **43**, 193-200.
43. Porter, H. v. K. Studies in the psychology of stuttering: XIV, Stuttering phenomena in relation to size and personnel of audience. *J. Speech Disor.,* 1939, **4**, 323-333.
44. Sheehan, J. G. A study of the phenomena of stuttering. M.A. thesis, University of Michigan, 1946.
45. ———. A theory of stuttering as approach-avoidance conflict. *Amer. Psychol.,* 1950, **5**, 469.
46. ———. The modification of stuttering through non-reinforcement. *J. Abn. & Soc. Psychol.,* 1951, **46**, 51-63.

47. ———. Fear-reduction during stuttering in relation to conflict, "anxiety-binding" and reinforcement. *Amer. Psychol.*, 1952, **7**, 530 (Abstract).

48. SHEEHAN, J. G., & VOAS, R. Tension patterns during stuttering in relation to conflict, fear-reduction, and reinforcement. *Speech Monog.*, (in press).

49. SHEEHAN, J. G., & ZELEN, S. A level of aspiration study of stutterers. *Amer. Psychol.*, 1951, **6**, 500 (Abstract).

50. SHULMAN, E. A study of certain factors influencing the variability of stuttering. Ph.D. dissertation, State University of Iowa, 1944.

51. SOLOMON, M. Stuttering as an emotional disorder. *Proc. Amer. Speech* Correct. Assoc., 1932, **2**, 118-121.

52. SOLOMON, M. Stuttering as an emotional and personality disorder. *J. Speech Disor.*, 1939, **4**, 347-357.

53. TRAVIS, L. E. The need for stuttering. *J. Speech Disor.*, 1940, **5**, 193-202.

54. ———. My present views on stuttering. *Western Speech*, 1946, **10**, 3-5.

55. TRAVIS, L. E., TUTTLE, W. W., & COWAN, D. W. A study of the heart rate during stuttering. *J. Speech Disor.*, 1936, **1**, 21-26.

56. VAN RIPER, C. A study of the thoracic breathing of stutterers during expectancy and occurrence of stuttering spasm. *J. Speech Disor.*, 1936, **1**, 61-72.

57. ———. Effect of devices for minimizing stuttering on the creation of symptoms. *J. Abn. & Soc. Psychol.*, 1937, **32**, 185-192.

58. ———. The effect of penalty upon frequency of stuttering. *J. Genet. Psychol.*, 1937, **50**, 193-195.

59. ———. The relation of tremors to perpetuation and release from the stuttering spasm. Unpublished paper read at 1939 meeting of the American Speech Correction Association, Chicago, Ill. See A.S.C.A. Program, *J. Speech Disor.*, 1939, 336.

60. ———. Speech Correction; Principles and Methods. (2nd ed.) New York: Prentice-Hall, 1947.

61. ———. Stuttering. Chicago: National Society for Crippled Children and Adults, 1949.

62. WISCHNER, G. J. Stuttering behavior and learning: A preliminary theoretical formulation. *J. Speech & Hear. Disor.*, 1950, **15**, 324-335.

63. WYNEKEN, C. Ueber das Stottern und dessen Heilung. (Concerning stuttering and its cure.) *Zeitschrift für Rationelle Medicin*, 1868, **31**, 1-29.

Department of Psychology
University of California
Los Angeles 24, California

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 51-57.

KINESTHETICALLY GUIDED MOVEMENTS OF HEAD AND ARM*

Department of Psychology, Kansas State College

EUGENE S. EDGINGTON[1]

A. PROBLEM

There is an obvious need for research on the rôle of the kinesthetic sense in motor performance. The present study was conducted for the purpose of obtaining knowledge concerning coördinated movements of head and arm; more specifically, the study was concerned with the ability of a person to look where he is pointing or to point where he is looking, when vision of his pointing hand and arm are excluded. The effects of the following factors upon the localizing responses were studied: (*a*) the angle of the head (in a horizontal plane) relative to body position, and (*b*) angular movement of the localizing organ during localization. Factor (*a*) was considered in both of the experiments performed, whereas the effect of Factor (*b*) was considered only in Experiment II.

B. APPARATUS AND PROCEDURE

The localization apparatus shown in Figure 1 was essentially a sheet of metal curved about and nailed to a piece of fiberboard. The metal is called the backplate in this paper. The backplate was so curved that all positions on the right half of the apparatus were equidistant from the right shoulder and all positions on the left half of the apparatus were equidistant from the left shoulder; therefore, the backplate was not curved in the form of a true semicircle.

A sight tube limited the subjects' peripheral vision to a 5½" section of the backplate. The sight tube was used only in the first experiment. Two pointers were used, one for the forefinger of each hand. Each pointer was a small thin piece of metal that was taped to the subject's forefinger so as to extend ½" beyond the end of the finger, in the center of the longitudinal axis of the finger.

---

*Received in the Editorial Office on January 24, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Now at Michigan State College.

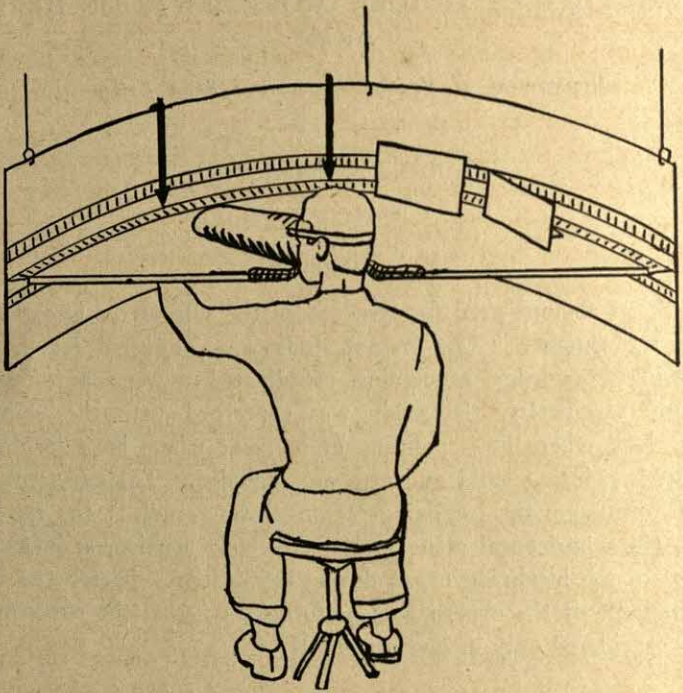


FIGURE 1
SUBJECT MAKING A LOCALIZING RESPONSE

Twenty subjects were used in Experiment I and 20 other subjects were used in Experiment II; no subject was used in both. The subjects were male college students between 18 and 30 years of age, with sufficient reaches to touch the backplate of the localization apparatus while seated at it. None of the subjects had any deformity which prevented normal movement of the head, arm, or hand; subjects with stiff necks, sore arms, or poor vision were not used. The subjects in both experiments were told only that the experiment was concerned with eye-hand coördination. They were not given any information about their own localizing responses or those of others until all of the data had been obtained from all 20 subjects used in the experiment.

The two sessions required for each experiment were scheduled to allow two to seven days between sessions. The subjects were asked not to practice making localizing responses or discuss the experiment with anyone during the interim. After the second session was completed, the subjects were again asked not to discuss the experiment with anyone.

The sequence of the localizing responses was randomized to nullify the effect of sequence upon the data. The numerical sequence in which the localizing responses were to be made was indicated in the margin of the data sheet to aid the experimenter.

The procedures for each experiment were the same for both sessions. Localizing responses were made with each arm; the right arm was used for localizing responses on the right half of the apparatus (relative to the subject seated at the apparatus) and the left arm was used for localizing responses on the left half. Localizing responses were made toward comparable positions for the right and left sides, making a localizing response with only one arm at a time, however. In both Experiments I and II the subject's body faced the front of the localization apparatus at all times; the body never turned. Also, in both experiments, during localizing operations, the direction in which the subject's head was turned coincided approximately with his line of sight. This control of the relationship of the eye axes to head position was accomplished through the use of the eye tube in Experiment I and by means of instructions and close observation by the experimenter in Experiment II. In either experiment, when the experimenter moved the subject's hand to the backplate, the subject's arm and hand were relaxed until his forefinger was placed against the backplate, at which time the subject tensed his arm and hand enough to keep his finger against the backplate.

For both Experiments I and II, each subject at the conclusion of his second session made an estimate of his maximum error for a single localizing response, taking all localizing responses for both sessions into consideration.

1. *Procedure for Experiment I*

The positions toward which localizing responses were made in Experiment I were *A, B, C, D, E,* and *F*. Position *A* was near the outer edge of the localization apparatus, on either side, and Position *F* was squarely in the middle of the backplate, separating the right half from the left half. Considering Position *F* to be 90°, Positions *A, B, C, D,* and *E* were approximately 5°, 20°, 40°, 55°, and 75°, respectively, relative to the notch in which the head rested.

The purpose of Experiment I was to determine the effect of head position in relation to body position on the magnitude, variability, and direction of errors made in making localizing responses with the head and with the arm. The subject started each localizing response from the same starting places: hands in the lap, with head facing the front, looking straight in front of his body at a reference marker.

a. Arm-to-Head. In the Arm-to-Head localizing response, the subject was told which direction to turn his head until he saw a marker, then to point to the marker by moving his hand straight away from his body beneath the partition until he touched the backplate with his pointing finger.

b. Head-to-Arm. In the Head-to-Arm localizing operation, the subject's relaxed hand and arm were moved by the experimenter until the subject's forefinger touched the backplate. The subject then held his arm tense in that position while he turned his head to look in the direction of his hand. The subject read combinations of letters from the backplate to indicate where he thought his hand was.

Cardboard sight blocks shown in Figure 1 were between the subject and the backplate, except where markers (arrows) were placed. (Figure 1 shows only two sight blocks in order to give the reader a better view of the apparatus.) A sight block in the line of sight of the subject was lifted for the subject to read the letters for the Head-to-Arm operation.

2. *Procedure for Experiment II*

The positions toward which localizing responses were made in Experiment II were *C* and *D*, which were in the middle section of the right and the left halves of the apparatus.

The purpose of Experiment II was to determine the effect of angular movement of the localizing organ during localization, with the head turned at various angles in relation to the body. The effect of the direction of angular movement (swinging) was also studied in this experiment. The sight tube was not used in this experiment because the effect of its weight and restriction of vision upon the results obtained in Experiment I was unknown.

a. Swinging Arm Out. This is an Arm-to-Head localizing operation. The subject's hand was placed by the experimenter at Location *F*, immediately in front of the subject, while he looked at the marker. Then the subject pulled his hand away from the backplate and swung it freely through the air until he thought he was pointing to the marker, at which time he placed his finger against the backplate.

b. Swinging Arm In. In this operation, the relaxed hand was placed at Position *A* instead of Position *F;* otherwise the procedure for Swinging Arm In was the same for Swinging Arm Out. For both swinging arm operations the subject was allowed to swing his arm as slowly or as rapidly as he desired, but he was not permitted to reverse the direction of the swing; as soon as the arm quit swinging the subject placed his finger against the backplate.

c. Swinging Head Out. This is a Head-to-Arm localizing operation. The procedure was the same as for the Head-to-Arm operation in Experiment I.

d. Swinging Head In. In this operation the subject's relaxed hand was moved to the backplate while the subject's head was turned, looking at a marker at Position *A;* therefore, the subject had to swing his head inward to make his Head-to-Arm type of localizing response.

Sight blocks were used in Experiment II in the same manner that they were used in Experiment I.

C. Results

The direction of the errors, as well as the magnitude, was considered in the analysis of the data for both experiments. Positive and negative signs denoted the direction of the errors, and of course this led to different results than would have been obtained had only the magnitude of the errors been used in the analysis. A "mean error" in either Experiment I or II indicates a position on the backplate about which the errors tended to fall. In Table 1, negative signs indicate that the localizing responses tended to cluster about a position nearer the center of the backplate than the position they were trying to localize and positive signs indicate the opposite tendency: the errors were the result of localizing responses being too far out, toward the edge of the apparatus.

1. *Results for Experiment I*

Snedecor's *F* test showed that the difference in variability between Arm-to-Head and Head-to-Arm, taking into consideration angular position, person, and side (right or left), was significant beyond the .001 level of confidence. Head-to-Arm showed the greatest variability.

TABLE 1

Ratios of Arm-to-Head and Head-to-Arm Variances and Mean Errors for Arm-to-Head and Head-to-Arm

Position	F ratio of variances	Greater variance	Arm-to-Head mean error	Head-to-Arm mean error
A	3.35*	H-to-A	—1.4†	—1.1†
B	3.99**	H-to-A	— .3	—1.7†
C	3.92**	H-to-A	+ .1	—1.3†
D	1.58	H-to-A	— .1	— .6†
E	1.11	H-to-A	+ .8†	— .6†
F	1.28	A-to-H	— .1	— .1

*Significant at the .05 level.
**Significant at the .01 level.
†Significant directional error.
Note.—The unit of "mean error" reported in this table is an inch.

At the .05 level, there was no difference in the variability from position to position, for Arm-to-Head. For Head-to-Arm, the variances for the positions were significantly different at the .05 level, Position *C* showing the greatest variance, and *F* the least.

When questioned at the end of the second session, 19 of the 20 subjects underestimated the magnitude of the error for their poorest response.

The analysis of the data for Experiment I depended upon 1,440 localizing responses.

2. *Results for Experiment II*

The difference in variability between Swinging Arm and Swinging Head, taking into consideration the side, direction of swing, and the position, was significant at the .01 level. Swinging Head showed the greater variability.

As readings were taken for both Swinging Arm and Swinging Head, with both arms, at Positions *C* and *D*, with the localizing organ swinging inward (toward the center) and outward, there were 16 combinations of these elements. Of the 16 combinations, nine showed significant directional errors. Five of these nine combinations involved Swinging Arm, and the other four involved Swinging Head. These directional errors were all due to the localizing organ not swinging far enough in the direction it was swinging. The seven non-significant combinations each showed a tendency for the same type of directional errors just mentioned, but not great enough to be significant at the .05 level.

For the same side, same position, and same direction of swing, there was a significant difference between the mean error for Swinging Arm and Swinging Head in every instance. In every case there was a significant difference between the mean error for Swinging In and the mean error for Swinging Out, for the same localizing organ, same side, and same position. As mentioned previously, these "mean errors" are actually positions on the backplate, so a difference of mean errors here means the distance between two positions about which responses tend to fall.

The algebraic sum of the errors (the sign denoting the direction of error) for each of the following categories was calculated for each subject individually: Swinging Arm In, Swinging Arm Out, Swinging Head In, and Swinging Head Out. The results showed that every one of the 20 subjects tended to not swing the localizing organ far enough, in at least three of the four categories.

All 20 subjects underestimated the magnitude of the error for their poorest response, when questioned at the end of the second session.

The analysis of the data for Experiment II depended upon 1,280 localizing responses.

D. SUMMARY

1. Two experiments were conducted to determine the influence of certain factors upon the ability of a person to look where he is pointing or to point where he is looking, when vision of his pointing hand and arm are excluded.

2. Differences in variability and differences of mean errors were both taken into consideration in the statistical analysis of the data. Mean errors were computed, considering the direction of the errors as well as the magnitude.

3. The variability of localizing responses was found to be greater when a person looks where he is pointing than when he points where he is looking.

4. In turning to where they were pointing, subjects did not move their heads far enough. When the arm was moved angularly, analogous to head movement, and in a constant direction (no back-tracking) subjects did not move the arm far enough when they were trying to point where they were looking.

358 North Harrison
East Lansing, Michigan

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 59-65.

A STUDY OF MEDICAL STUDENTS WITH THE *MMPI*: I. SCALE NORMS AND PROFILE PATTERNS*

Department of Psychiatry and Neurology, University of Minnesota Medical School

WILLIAM SCHOFIELD

A. INTRODUCTION

For the past four years the writer has given a series of 10 lectures on principles of psychometric testing and representative techniques to the junior medical students of the University of Minnesota at the time of their psychiatric clerkship. Prior to each lecture series the junior clerks took the Minnesota Multiphasic Personality Inventory (6), using the booklet form for individual, self-administration. After several lectures on the *MMPI* the clerks received their profiles individually and were invited to raise questions of interpretation in class or privately. In casually examining the profiles of each clerkship group the writer was impressed with the frequency of profiles showing at least one deviant score (i.e., T-score $\geqslant 70$) on the clinical scales. This finding motivated a preliminary survey of the *MMPI* profiles of 201 male junior medical students. These profiles comprised a nearly 100 per cent sample of the male juniors for the years 1945-47. During this period there were only 34 female junior medics.

An analysis of the *MMPI* data for the large male sample revealed that over *one-fourth* of the junior clerks had a score of 70 or greater on at least one clinical variable. . Of the total number of elevated scores, 27 per cent were obtained on the Masculinity-femininity scale (*Mf*), indicating deviation in the direction of femininity. The second most frequently elevated scale was hypomania (*Ma*), with the Psychopathic Deviate (*Pd*) scale a close third. The median *Mf* score for the male medical juniors was approximately 60, indicating that 50 per cent of this group have scores on this scale of a magnitude expected in about 16 per cent of the general population.

On the basis of these findings it was decided to make a careful normative study of the *MMPI* with a specific group of medical students, attempting to establish: (*a*) frequency of deviant scores per scale and of specified profile patterns, (*b*) changes in scores and profile after two years of medical training, and (*c*) relationship between *MMPI* profiles and academic achievement in medical school. The freshman medical class of 1946 was selected for this study since it was the first freshman class to be administered the *MMPI* as a group; this testing took place following admission of the students to the medical school. Junior year profiles were also available for this group of students; this made possible a study of the individual and group changes on the *MMPI* variables as well as changes in the score patterns. The office of the Dean of the Medical School contributed data on academic achievement of the members of this class which permitted a study of the relationship of the *MMPI* variables to success in the first three years of the medical curriculum. This paper will report the results of a normative study of the *MMPI* data for this class of medical students with comparative data from other medical student samples.

---

*Received in the Editorial Office on January 27, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

B. Medical Student Norms

Table 1 presents measures of central tendency on the clinical scales of the *MMPI* for two groups of male, beginning medical students. The Wisconsin students were in the third year of their pre-medical curriculum and had not yet been formally admitted to the medical school.[1] In view of their uncertain status as regards medical school admission, it would be reasonable to expect a degree of defensiveness which might make for some invalidity of their clinical scores. Actually, the scores of this sample on the validity measures (?, *L*, and *F*) were within normal limits, and their average score on the *K*-scale was 54.4 which is well below the median *K*-value commonly reported for college persons (2). The Minnesota sample consisted of students who were tested subsequent to their admission to medical school, at the beginning of their freshman year. They did not have the motives to defensiveness possibly operant for the Wisconsin students and their profiles might be expected to be more directly reflective of any personality trends specific to such a group. The Minnesota validity scales were within normal limits; the mean *K*-score was 59.9, which is slightly higher than that of the Wisconsin sample but closely similar to that reported for various college samples (2).

The mean scores of the two freshmen groups, as reported in Table 1, are very similar on eight of the nine clinical *MMPI* scales; the average difference between comparable means on the first eight scales is only 1.7 *T*-score points. On the *Ma* scale, the Minnesota group produced a mean nearly one sigma higher than that of the Wisconsin students. The essential

---

[1] The data for this sample were generously provided by Dr. Lewis E. Drake of the University of Wisconsin Student Counseling Center.

TABLE 1

MEAN SCORES ON THE *MMPI* CLINICAL SCALES FOR TWO SAMPLES OF MALE MEDICAL STUDENTS AND CODES* FOR THE MEAN PROFILES

Sample	N	?	L	F	K	Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Wisconsin 3rd year pre-medic	68	50	50	50	54.4	46.7	52.8	55.7	50.0	60.8	52.7	48.1	47.4	50.8
			K-corrected:			48.0			52.7			49.6	50.0	51.8
Minnesota freshman	83	50	50	50	59.9	45.8	49.4	54.4	49.4	59.4	50.0	45.8	46.0	57.0
			K-corrected:			50.9			54.1			52.8	53.5	58.7

Codes: Wisconsin '53— (non-K).
Minnesota '593— (non-K).

---

*In brief, an *MMPI* code is a method for the quick numerical specification of the essential conformation and elevation of a profile by recording the number of the scales ($Hs = 1$, $D = 2$, etc.) in order of magnitude of the accompanying scores (4).

similarity of the two mean profiles is further reflected in the respective codes which indicate major deviations in both to occur on the *Mf* and *Hy* scales, in that order, with *Ma* as an intermediate high point in the Minnesota profile. In both mean profiles, $Hy > D > Hs$. *Hs* has the lowest mean score in both samples.

Table 2 indicates the frequency with which each of the clinical scales is the high point in the profiles of a sample of general population males, a general sample of University of Wisconsin male students, the sample of Wisconsin male pre-medics, and the sample of Minnesota freshman medical students. The general student sample differs chiefly from the general population sample in revealing a markedly greater frequency of high points on *D, Hy, Pt, Sc,*and *Ma,* and markedly smaller frequency of high points on *Hs.* All of these differences between proportions yield critical ratios exceeding the 1 per cent level of probability. The relative frequency of high points in these two samples of normals is reliably different for six of the eight scales for which the comparison may be made. Also, the student sample showed reliably smaller frequencies of "indeterminate" profiles and of profiles with no high point. *A profile is coded as "indeterminate" if the three highest scale values are equal or within one point of each other.* As coded, such a profile yields a code in which the first three digits would be

TABLE 2
THE RELATIVE FREQUENCY WITH WHICH EACH OF THE *MMPI* SCALES OCCURRED AS THE HIGH POINT IN THE PROFILES OF FOUR SAMPLES OF MALES†

Code	Scale	Atlas normal males $N=256$	Univ. of Wisc. students $N=2805$	Wisconsin pre-medics $N=68$	Minnesota Freshman medics $N=83$
1	Hs	12.9%	3.0**%	0.0**%	4.8%
2	D	8.2	13.9**	16.9	2.4**
3	Hy	7.1	15.4**	14.7	7.2
4	Pd	11.3	8.5	2.2**	5.4
5	Mf			36.7	32.5
6	Pa	8.6	5.1	5.8	1.8
7	Pt	6.7	13.4**	5.8**	6.0
8	Sc	4.8	9.0**	4.4	4.2
9	Ma	18.3	25.6**	11.8**	22.3
Indeterminate‡		23.6	5.6**	1.5**	10.9*
No high-point#		23.7	6.3**	0.0**	2.5

†Asterisks in any column denote a statistically reliable difference between the starred values in that column and the comparable values in the *preceding* column.
*Significant at the 5 per cent level.
**Significant at the 1 per cent level.
‡See text.
#See text.

underlined. Indeterminacy pertains to the fact that it would not be possible to index such a code under a single, two-digit index as such indices were defined in compiling the frequency charts of the *Atlas* (5). *A profile with "no high point" is one in which no one of the clinical scores exceeds* $T = 54$*.*

When the high-point frequencies of the Wisconsin pre-medical students are compared with those for the general Wisconsin student sample, reliable differences are found on the *Hs, Pd, Pt,* and *Ma* scales. The medical student sample shows a significantly smaller frequency of high points than the general student sample on each of these scales. Indeterminate profiles and profiles with no high point also had reliably greater frequency in the general student sample.

A comparison of the frequency of high points in the two medical student samples reveals a reliable difference on the Depression scale only, the Wisconsin group showing a significantly greater occurrence of high points on this scale than found for the Minnesota group.

In summary, Table 2 indicates that college males in general show considerable difference from general population males in the frequency of their high points, that two samples of male medical students differ more from a male general college sample than they do from each other, and that both samples of medical students are characterized by having their most frequent high point on the *Mf* scale.

Table 3 indicates the per cent of the Minnesota freshman medics and Wisconsin pre-medics whose scores exceeded $T = 70$ on each of the clinical scales. Normal expectancy for such scores, because of the slightly skewed distributions obtained with the scales, would be from 3 to 5 per cent. The similarities between the Minnesota freshman medical students and the Wisconsin pre-medics is particularly notable in the absence of deviant (>70) scores on the *Hs* and *Pa* scales; the low frequency of deviant scores on *Hy*, *Pd*, and *Sc*; and the considerable frequency of deviations on the *Mf* scale.

TABLE 3

PER CENT OF MINNESOTA AND WISCONSIN MALE MEDICAL STUDENTS WHOSE SCORES EXCEEDED $T = 70$ ON EACH OF THE *MMPI* CLINICAL SCALES

Group		Hs	D	Hy	Pd	Mf	Pa	Pt	Sc	Ma
Minn. Freshmen $N=83$	Per cent of scores >70	0	0	0	1.2	15.7	0	4.8	1.2	9.6
Wisc. $N=68$		0	4.4	1.5	2.9	16.2	0	7.3	2.9	0

The *D* and *Ma* scales are the only ones on which the two samples appear quite different, the differences being such as to suggest consistently a greater incidence of generally depressive, pessimistic, insecure orientation in the Wisconsin students. At least to some degree, this difference between the personality trends of the Wisconsin and Minnesota samples may reflect the difference in status of the two, the former being still anticipatory of, and undoubtedly anxious about, admission to medical school, whereas the Minnesotans had passed this hurdle and were tested when perhaps still experiencing the flush of enthusiasm at entrance upon their medical studies proper.

The complete bibliography of *MMPI* research studies published through December, 1949, was examined for studies on college populations. Only a few studies reporting summary statistics for sizeable samples of undergraduate males were found (1, 2, 3). These studies yielded mean profiles showing most of the characteristics of the mean profiles of the medical student groups as described above.

C. SUMMARY AND CONCLUSIONS

This paper reports normative data on the *MMPI* for two samples of undergraduate male medical students and compares the frequency of dominant profile characteristics of these samples with comparable frequencies for general population and college samples reported in the *Atlas* (5).

The medical student samples were comprised of 83 males in the class of 1950 at the University of Minnesota whose freshman *MMPI* records were obtained after their admission to medical school, and 68 male premedics at the University of Wisconsin tested in their last pre-medical year.

Mean profiles are reported for these two samples and the relative frequency with which each of the clinical *MMPI* scales appeared as the high-point in individual profiles is reported for the two medical samples and compared with similar data for a sample of normal adult males and for a large sample of college students. The following conclusions are suggested:

1. It was found that the two medical student samples showed a high degree of similarity in their *MMPI* profiles. The mean profiles of the two samples were characterized by primary elevations on the *Mf*-scale with secondary peaks on the *Hy*-scale. The only scale revealing a reliable difference in the mean scores of the two groups was *Ma,* the Minnesota sample showing a higher mean score on this scale.

2. When the frequency with which each of the clinical scales yielded the high-point in profiles of the two medical student samples was studied, only one reliable difference appeared. The Minnesota sample showed a re-

liably smaller frequency of high-points on the *D*-scale. This difference suggests that the morale of the Minnesota students, who had already been accepted into the medical school, was superior to that of the Wisconsin students who were still at the pre-medical level. The markedly greater frequency of high-points on *Ma* in the Minnesota sample, while not statistically reliable, supports this hypothesis.

3. In terms of high-point frequencies, the two medical student groups were found to be more like each other than like a general male college sample. The general male college sample showed a distribution of high-point frequencies which differed markedly from that of a sample of non-college, normal males.

4. A comparison of the mean *MMPI* profiles of the medical students with those of other samples of male college students did not suggest any basic personality pattern distinctive for the medical group.

REFERENCES

1. BLUM, I. P. A comparative study of students preparing for five selected professions including teaching. *J. Exp. Educ.*, 1947, **16**, 31-65.
2. BROWN, H. S. Similarities and differences in college populations on the Multiphasic. *J. Appl. Psychol.*, 1948, **32**, 541-549.
3. GILLILAND, A. R., & COLGIN, R. Norms, reliability, and forms of the *MMPI*. *J. Consult. Psychol.*, 1951, **15**, 435-438.
4. HATHAWAY, S. R. A coding system for *MMPI* profile classification. *J. Consult. Psychol.*, 1947, **11**, 334-337.
5. HATHAWAY, S. R., & MEEHL, P. E. An Atlas for the Clinical Use of the *MMPI*. Minneapolis: Univ. Minnesota Press, 1951.
6. HATHAWAY, S. R., & MCKINLEY, J. C. The Minnesota Multiphasic Personality Inventory. Minneapolis: Univ. Minneapolis Press, 1943.

Department of Psychiatry and Neurology
The Medical School
University of Minnesota
Minneapolis 14, Minnesota

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 67-81.

THE EFFECTS OF REINFORCEMENT ON THE OCCURRENCE OF CLUSTERING IN THE RECALL OF RANDOMLY ARRANGED ASSOCIATES[*][1]

Department of Psychology, University of Connecticut

W. A. BOUSFIELD AND B. H. COHEN

A. THE PROBLEM

Clustering is here defined as the occurrence of sequences of related words as they appear in the recall of items presented for learning in random order. The purpose of this study is to investigate the influence on this phenomenon of varying numbers of presentations of a list of stimulus-words. The theoretical significance of clustering derives in part from the tenable assumption that it represents a type of organization characteristic of higher order mediating processes. We may paraphrase Hebb (2) by stating that such processes are not stimulus-bound in the sense of being under full and direct stimulus-control. At the same time they do coöperate with identifiable aspects of stimulation. By studying the characteristics of this coöperation, we hope to learn something of the nature of mediating processes and the laws under which these processes operate.

The methods for inducing and quantifying clustering employed in this study are similar to those used by the senior author (1) in an earlier investigation. His technique involved the use of a list of 60 nouns comprising 15-each of four separate categories, namely, *animals, names, professions,* and *vegetables.* These nouns were presented verbally and in random order to groups of subjects who, following the presentation, wrote in order as many nouns as they were able to recall within a period of 10 minutes. In analyzing the data, any sequence of two or more items in the same category was regarded as a cluster. The results of this study indicated clustering appreciably beyond chance expectation. Furthermore, the clustering tendency changed in an orderly manner from the beginning to the end of recall. It was initially somewhat above chance, rose to a maximum, and then declined to apparent randomness. The present study differs from the early one in two essential respects. Instead of using only one reinforcement of the stimulus-list, the

---

[*] Received in the Editorial Office on January 28, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This is Technical Report No. 1 under contract Nonr—631 (00) between the Office of Naval Research and the University of Connecticut.

words were given varying numbers of reinforcements ranging from one to five. Secondly, we shifted from a verbal to a visual presentation of the words.

In order to clarify the purpose of this study, we indicate below the hypotheses we proposed to test:

1. The total amount of clustering occurring in the recall of randomized verbal associates should be a positive function of the degree of their reinforcement.
2. The progressive changes in clustering from the beginning to the end of recall of a group of randomized associates should be modified by reinforcement in the following ways:
    a. The initial level of clustering should be a positive function of reinforcement.
    b. There should be a positive relationship between reinforcement and the speed of attainment of maximum clustering.
    c. For all levels of reinforcement clustering should tend to decrease as the supply of available associates approaches exhaustion.

Basically these hypotheses derive from two assumptions. The first of these is that the organizing tendency responsible for clustering is developed by learning occurring during both reinforcement and recall. The second is that the effectiveness of the organizing tendency becomes reduced as the supply of organizable material approaches exhaustion. We shall reserve the consideration of the theoretical bases of these assumptions for later discussion in this paper.

B. METHOD

1. *Subjects*

The subjects were 243 undergraduate students enrolled in 10 laboratory sections in introductory psychology. These sections were selected on the basis of approximately equivalent achievement in their course examinations at the middle of the semester. Their sizes ranged from 18 to 31 students. The 10 sections were in turn combined at random into five groups, each comprising the students in two sections. The sizes of these groups were as follows: Group I, 49 subjects; Group II, 52 subjects; Group III, 46 subjects; Group IV, 49 subjects; Group V, 47 subjects.

2. *Stimulus-Materials and Apparatus*

We employed the stimulus-word-list of the earlier experiment (1). As already noted, this list comprised 60 two-syllable nouns with a 15-each in the following four categories: *animals, names, professions,* and *vegetables.* In order to equate the categories as closely as possible with respect to the associative values of their items, use had been made of the Thorndike-Lorge (3) frequency tables. These words with their frequencies per million appear in Table 1.

TABLE 1
LIST OF STIMULUS-WORDS

Animals		Names		Professions		Vegetables	
Word	Freq.	Word	Freq.	Word	Freq.	Word	Freq.
giraffe	1	Amos	1	milkman	1	eggplant	1
baboon	2	Gerald	1	typist	1	parsnip	1
zebra	2	Byron	3	florist	2	garlic	3
panther	5	Oswald	4	plumber	2	rhubarb	3
wildcat	5	Jason	5	diver	4	radish	4
leopard	6	Otto	5	druggist	4	melon	5
reindeer	6	Noah	6	broker	7	mustard	7
chipmunk	7	Wallace	7	printer	7	spinach	8
muskrat	7	Owen	7	dentist	9	parsley	8
woodchuck	7	Bernard	9	baker	10	carrot	9
otter	8	Adam	10	chemist	10	mushroom	10
weasel	9	Sherman	10	dancer	10	turnip	10
badger	11	Simon	11	grocer	11	lettuce	12
donkey	16	Moses	12	waiter	13	pumpkin	13
camel	18	Howard	19	blacksmith	19	cabbage	16
Mean Freq.	7.33		7.33		7.33		7.33

Twenty randomizations of this list were prepared on the basis of the drawing without replacement of 60 capsules from a box. From the 20 randomizations thus prepared, those lists containing runs (clusters) of five or more items in the same category were rejected. The remaining lists were then arranged in the order of the degree to which their total clustering approximated theoretical chance expectation. From these in turn we selected the five lists most closely approximating the theoretical norm. Such precautions were taken in order to insure optimal randomness of clustering in the stimulus-word-lists and thus avoid giving undue emphasis to the categories. A set of lantern slides was then prepared for each of the five randomizations of the words. To expose the words on a screen we used a projector with a mask which could be moved from word to word by the experimenter and in time with the flash of a small light occurring at three-second intervals. The light was mounted on the projector but could not be seen by the subjects. Mimeographed data sheets were prepared for use by the subjects in writing the words they were able to recall. These sheets were 8½" x 11" in size and divided by vertical lines into five columns numbered in order.

3. *Procedure*

The general procedure was as follows: the subjects in Group I received a single presentation of the stimulus-words; Group II received two successive presentations; Group III, three presentations; Group IV, four presentations; Group V, five presentations. Since five randomizations of the stimulus-words had been prepared, it was possible to avoid repetition of a serial order in all cases of multiple presentations of the list. This precaution appeared desirable in view of the assumption that memorization of a serial order of the words would operate against the organizing tendency under investigation.

The experiment was conducted in the small lecture room where the students regularly assembled during the scheduled period, the apparatus having been set up in advance. In conducting the experiment, the experimenter distributed the mimeographed data sheets and delivered a brief prepared statement requesting coöperation in the undertaking. He then gave instructions as follows:

> Will you kindly print your name and section number in the top right hand corner of the paper which has been given you.
>
> By means of the lantern which you see in the middle of the room, a list of words will be projected, one after another on the screen in front of you. After completion of the projection, you are to recall as many words as you can, and write them in the order in which they occur to you. You will start by writing the words in Column 1 of the data sheet you have received. I shall give the signal to start writing by saying, "Ready to write. Go!" You will continue writing in the first column until I say, "Shift to the second column!" On hearing this signal, shift to the second column and continue writing the words as you recall them. This shifting will take place for each of the columns appearing on your data sheet. Please remember you are not to write any words until you are told to do so. Are there any questions?

After answering such questions as appeared relevant, the experimenter went to the projector and continued the instructions:

> Will you give your attention to the lighted area on the screen where the words will be shown. Remember you are not to write any words until the signal to write is given. Ready for the words. Now!

The words were then projected at the rate of three seconds per word. For the multiple presentations there was a pause of 30 seconds between each presentation. In every case the experimenter gave the signal to start writing the words recalled three seconds after the showing of the final word of the final presentation. The instructions to shift to the next column and continue with more words were given at two-minute intervals. In this way

10 minutes were allowed for recall, this time being divided into five two-minute intervals, one for each of the five columns on the data sheets.

C. Results

We shall summarize the analysis of our data having a bearing on the purpose of this report. Several other analyses, such as those dealing with the relationship between time spent in recall and the number of items produced, will be handled in other reports.

1. *Preparation of Data*

In order to facilitate the handling of the data, several steps were taken in preparing the lists of recalled words. First, the words on each data sheet were numbered in serial order. Each word was then labelled according to its classification as follows: *A,* for *animals; N,* for *names; P,* for *professions; V,* for *vegetables; I,* for words which were illegible, or if legible did not belong to any of the four categories of the stimulus-list. In addition, the symbol *NL* was added for words belonging in the *A, N, P,* or *V* categories but not appearing in the stimulus-list. Thus two types of errors were identified: (*a*) the *I*-words, which may be termed irrelevant intrusions; (*b*) the *NL*-words which may be termed categorical intrusions. Having classified the words in this way, all sequences of two or more words in the same category (including the *NL*-words) were bracketed to indicate the incidence of clustering. It may be noted that our designation of clustering makes no distinction between "correct" responses and categorical intrusions. This follows from our operational definition of a cluster as a sequence of words identifiable as belonging to the same category.

2. *Relationship between Reinforcement and the Totals of Words Recalled*

We undertook this treatment of the data in order to obtain an independent measure of the potency of reinforcement as here manipulated. The data from this analysis appear in Table 2. These data show the expected increase

TABLE 2
Relationship between Reinforcement and Total Recall

Reinforcements	No. of subjects	Mean no. items recalled per subject	Standard deviation of items recalled per subject
1	49	23.95	6.96
2	52	29.12	6.30
3	46	31.70	6.23
4	49	35.12	7.00
5	47	37.87	7.76

in the number of items recalled as a function of reinforcement. The progression of the means apparently proceeds at a negatively accelerated rate. It is obvious that we have covered only a limited range of possible reinforcement.

3. *Relationship between Reinforcement and Intrusions Appearing in Recall*

We have defined two types of errors, or what we prefer to call intrusions, appearing in the lists of recalled items. Data summarizing the incidence of these two types of intrusions appear in Table 3. It appears from these data that there were appreciably more categorical intrusions than those of the irrelevant type. The incidence of both types, however, is relatively small. Furthermore, while reinforcement is effective in reducing the categorical intrusions, it seems that the irrelevant intrusions are not influenced in any consistent way by reinforcement. We are inclined to regard the irrelevant intrusions as being primarily the result of tangential associative processes of sufficient potency to instigate the response of recording them. Roughly 25 per cent of them were indecipherable words which were classed as irrelevant because they could not be classed in any of the experimental categories. The number of these intrusions, however, is too small to have a material effect on the analyses of the data.

TABLE 3
RELATIONSHIP BETWEEN REINFORCEMENT AND THE INCIDENCE OF INTRUSIONS

Reinforcements	No. of subjects	Mean no. of categorical intrusions per subject	Mean no. of irrelevant intrusions per subject
1	49	1.41	0.18
2	52	0.73	0.21
3	46	0.39	0.26
4	49	0.55	0.10
5	47	0.36	0.30

4. *Relationship between Reinforcement and Total Clustering as Measured by an Index of Repetition*

This was the first of two analyses undertaken in order to test directly the first of our experimental hypotheses stating that the total amount of clustering should be a positive function of reinforcement. The method was the same as that used in the earlier study of clustering (1). It involved the use of a formula, derived from the binomial theorem, to determine the extent to which the subjects exceeded chance in their tendency to produce sequences of items in the same category. This formula provided an index of repetition, *IR*, which was applied to each of the lists of recalled items. Its statement is as follows for our design involving four possible categories for each item:

$$IR = \frac{4r}{\sqrt{3n}} - \sqrt{\frac{n}{3}} \tag{1}$$

The term r of this formula represents the total number of repetitions of words in similar categories. The term n is the total number of sequences, actually one less than the total number of items on a list. The *IR* has the advantage of being based on the assumption of a distribution with unit standard deviation about a mean of zero, and is independent of the size of n. Application of this formula to the data for each degree of reinforcement gave the results shown in Table 4. In spite of the deviant value of the *IR* for three reinforcements, it appears that the marked trend of the data supports the experimental hypothesis. It is obvious from the standard deviations that there was considerable variability in the data. Because of this fact, we hesitate to assume generality from these data beyond the trend which was predicted from the experimental hypothesis.

TABLE 4

RELATIONSHIP BETWEEN REINFORCEMENT AND GENERAL CLUSTERING AS MEASURED BY THE INDEX OF REPETITION

Reinforcements	No. of subjects	Mean index of repetition	Standard-deviation of indices of repetition
1	49	2.26	1.56
2	52	2.43	1.62
3	46	3.87	2.32
4	49	3.69	2.29
5	47	4.78	2.06

5. *Relationship between Reinforcement and the Number of Items Occurring in Clusters of Various Size*

If general clustering is positively related to reinforcement, it would also follow that with increasing amounts of reinforcement, we would expect an increasing incidence of items in the higher-order clusters. This assumption provides the basis for an additional test of the relationship between reinforcement and general clustering. To make this analysis a tabulation was made for each level of reinforcement of the number of items occurring singly and in clusters of varying size. To compare these data with those deriving from chance expectation we made a similar analysis of clusters found in the data of the artificial experiment reported in the earlier study (1). This experiment was run as an empirical check on the validity of the formula for the index of repetition and involved the drawing of capsules in four colors for 100 hypothetical subjects. Table 5 shows the mean number of items occurring singly and in clusters of varying size both for the subjects and for chance expectation as derived from the artificial experiment. The data representing chance have been adjusted for the changes in the total number of items recalled for each level of reinforcement. We may observe that the largest clusters in the artificial experiment comprised six items. The subjects, however, attained clusters of 12 items. These results show that with increase in reinforcement there tends to be an increase in the number of items in the higher-order clusters and thus a progressive departure from clustering occurring on the basis of chance. It may be noted that while our theory would predict a larger number of single (unclustered) items for chance than for the subjects, the differences for clusters of two items tend also to occur in the same direction. This raises the interesting question of the significance of the two-item clusters.

An attempt was made to find a suitable equation to describe the relationship between the size of the clusters and the mean number of items they comprised for each degree of reinforcement. Among the several possibilities we tried, the following exponential equation fitted the data fairly closely:

$$y = ae^{-bx} \tag{2}$$

In this function, y represents the mean number of items for the size of the cluster x. The other terms are a, the intercept of the curve on the vertical axis; e, the base of natural logarithms; and b, the slope factor. Since we cannot at present satisfactorily rationalize this equation, and other functions may be found to fit the data better, it appears inadvisable to claim significance beyond the apparent fact that the data appear amenable to this type of exploration.

6. *Relationship between Reinforcement and Progressive Changes in Clustering During Recall*

Our experimental hypotheses made three predictions regarding the effects of reinforcement on the nature of the progressive changes in the clustering tendency during recall. With increasing reinforcement the initial level of clustering should be raised and maximum clustering should be attained more rapidly. Furthermore, the clustering for all levels of reinforcement should show a tendency to drop in the direction of randomness as the supply of avail-

TABLE 5

RELATIONSHIP BETWEEN REINFORCEMENT AND THE MEAN NUMBER OF ITEMS OCCURRING IN CLUSTERS OF VARYING SIZE*

Rein-forcements		Single items	Size of clusters										
			2	3	4	5	6	7	8	9	10	11	12
1	Subjects	7.53	6.08	4.84	3.02	1.22	0.98	0.29	—	—	—	—	—
	Chance	13.93	6.58	2.50	0.69	0.19	0.06	—	—	—	—	—	—
2	Subjects	9.29	8.08	4.62	3.00	1.44	1.50	0.54	0.31	0.35	—	—	—
	Chance	16.94	8.00	3.05	0.84	0.23	0.07	—	—	—	—	—	—
3	Subjects	7.61	6.26	3.98	3.13	3.04	1.96	2.13	1.57	1.37	0.65	—	—
	Chance	18.43	8.71	3.31	0.91	0.25	0.08	—	—	—	—	—	—
4	Subjects	9.37	7.96	4.29	2.61	2.76	2.08	1.71	1.63	1.65	0.82	0.00	0.24
	Chance	20.42	9.65	3.67	1.00	0.28	0.08	—	—	—	—	—	—
5	Subjects	8.21	6.46	4.53	3.74	2.23	3.83	2.83	2.72	0.96	0.64	0.94	0.77
	Chance	22.03	10.41	3.96	1.09	0.30	0.09	—	—	—	—	—	—

*The minimal size of a cluster is two items. The means in the first column are therefore for single (unclustered) items.

able items approaches exhaustion. The testing of these predictions necessitated the devising of a valid index to represent the changes in the amount of clustering occurring from the beginning to the end of the lists of recalled items. The first consideration arising in this undertaking was that of dividing the lists of recalled items into comparable units. Obviously the lengths of these lists varied over a wide range. To accomplish this we used the Vincent-method in dividing each list into successive decile-intervals. For an index to represent the magnitude of clustering we first considered employing the *ratio of repetition* which was used in the earlier study (1). This is the decimal fraction obtained by dividing the number of repetitions by the total number of items. After comparing this index with several other possibilities we abandoned it in favor of what we may call the *density ratio*. This new and simple index may be defined as the decimal fraction obtained by dividing the number of clustered items by the total number of items. The basic difference between the old and the new indices is apparent in their designation of clusters. A cluster of four items, for example, would have three repetitions and four clustered items. The density ratio is thus larger than the ratio of repetition. Applying the two indices to the same samples of lists of items we found empirically that both yielded essentially the same profile for progressive changes, but the density ratio was simpler to compute and produced a smoother curve.

The density ratios per decile for each of the five levels of reinforcement are shown in Table 6. Smoothed curves of these data appear in Figure 1.

TABLE 6
RELATIONSHIP BETWEEN REINFORCEMENT AND DENSITY RATIOS PER DECILE OF WORDS RECALLED

	Reinforcements				
Decile	1	2	3	4	5
1	.602	.631	.799	.809	.859
2	.713	.769	.806	.852	.875
3	.839	.779	.894	.839	.863
4	.752	.772	.842	.831	.861
5	.798	.793	.880	.800	.843
6	.849	.786	.840	.747	.821
7	.615	.621	.773	.750	.811
8	.600	.592	.718	.637	.725
9	.538	.548	.650	.584	.697
10	.522	.522	.442	.500	.484

The smoothing was done by computing a running average of three deciles, keeping unchanged, however, the ratios for the first and tenth deciles. A reference line representing chance expectation also appears in Figure 1. This was derived from a parallel analysis of the data from the artificial experiment, to which reference has already been made, and represents a mean density ratio of .417.

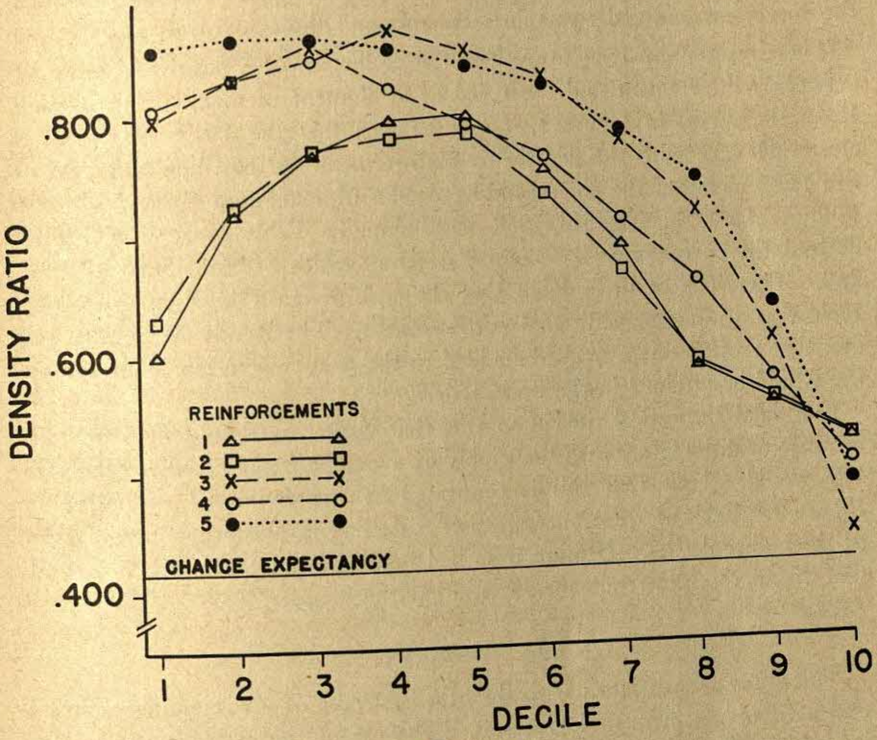


FIGURE 1
PROGRESSIVE CHANGES IN CLUSTERING DURING RECALL AS A FUNCTION OF REINFORCEMENT

Inspection of these curves in Figure 1 reveals noticeable variability which is especially apparent in the curve for three reinforcements. The discrepancy in clustering for three reinforcements has already been noted in the data of Table 4 where the mean index of repetition is higher for three reinforcements than it is for four. Paradoxically this inversion did not appear for the mean number of items recalled as shown in Table 2. In this connection it may be noted that though general clustering, as measured by the index of repetition, correlates positively with the total items recalled, our correlations have never exceeded .38. The variance which rote memory contributes to clustering is evidently small. We are at present inclined to suppose that the variability of our data reflects uncontrolled peculiarities in our populations of subjects.

It is evident that we must depend on trends in the interpretation of our data. These trends, however, are in support of our experimental hypotheses. The initial densities of clustering tend to vary positively with degree of reinforcement. This is especially apparent if a gross comparison is made between the curves for one and two reinforcements and those for 3, 4, and 5.

These results also tend to confirm the prediction of a positive relationship between reinforcement and the speed of attainment of maximum clustering. From the curves in Figure 1, it may be seen that the deciles of maximum clustering vary with the degree of reinforcement. Proceeding from one to five reinforcements the corresponding deciles of maximum clustering are as follows: 5th, 5th, 4th, 3rd, 2nd. The data in Table 6, from which we derived the curves, give the slightly different order of: 6th, 5th, 3rd, 2nd, 2nd. There can be little doubt that the peak of clustering is reached more rapidly as a result of additional reinforcement. The trends of the data suggest that reinforcement might be added to the point where maximum clustering would be attained in the first decile.

Our final theoretical prediction was that clustering should undergo some measure of disintegration as the supply of available items becomes exhausted. We would submit that this is confirmed by the trends of the curves in Figure 1. Each of these curves shows a decline in clustering in the region of the higher deciles. If more time had been allowed for recall, it is likely that the approach to randomness would have been closer.

D. Discussion

This study was designed with the broad purpose of investigating organization as it occurs in mental processes. The use of the experimental method in such an undertaking must of necessity deal with rather specific aspects of the problem and these aspects in turn must be definable in terms of experimentally manageable operations. We have assumed that the clustering occurring in the recall of words is a consequence of organization. By observing the effects of reinforcement on clustering, we are provided with information useful in developing a theoretical conception of the nature of organization as it occurs in mental processes. We shall undertake here to indicate the outlines of a theoretical account of the mechanisms responsible for clustering, and to relate this theory to the experimental hypotheses which we believe are supported by the results of our study.

The basic problem in developing this theoretical analysis is that of accounting for the occurrence of individual items during recall. In the earlier study (1) it was assumed that the probability of occurrence of any single item would be contingent on the magnitudes of two types of strength which the item would possess as a potential verbal response. One of these would be habit strength deriving from the various types of reinforcement the word has received both before and during the experiment. The second type of strength was termed *relatedness increment;* a word by virtue of its occurrence adds this increment to other words to which it is related. Thus if a subject names an animal, a relatedness increment is added to other animal-words. The words are related by virtue of belonging to the same category. We would propose to develop these assumptions on the basis of Hebb's (2) account of the development of superordinate perceptions. In this connection, Hebb proposes that we are dealing with a learning situation in which the repeated perception of related parts results in the emergence of a new structure. the activity of which constitutes the superordinate perception. Eventually the activation of a single perceptual element may be sufficient to excite the superordinate system which, in turn, facilitates the responses of its subordinate systems. Following this line of reasoning, we may suppose that the production of a single word in our type of experiment will tend to activate its superordinate system, namely that which represents the category to which the word belongs. This superordinate system would then tend to facilitate its subordinate word-habits. What we have termed the relatedness increment would thus result from the facilitating activity of the superordinate systems. A shift in categories would occur during recall at any time as a result of sufficient strength of a word in a different category. It is important to note that according to these conceptions, the superordinate systems would become strengthened as a result of repeated presentations of the list of stimulus-words during the learning periods. This would account for the positive relationship between reinforcement and: (*a*) total clustering; (*b*) the initial level of clustering. It would further follow, however, that the superordinate systems would tend to acquire additional strength because of their activation during recall. It is not unreasonable to assume that the hypothetical learning curve of a superordinate system would be of the negatively accelerated type. At the time of recall, a superordinate system that is initially strong is nearer its maximum strength than one that is initially weak. The strong system would therefore gain less from reinforcement during recall than the weak one. The limiting case would be a superordinate system that had attained maximum strength before the beginning of recall. From these assumptions it would follow that the possible range of development of superordinate system during recall would depend on their initial strengths. As a result of reinforcement during recall the initially weak superordinate systems would continue gaining relatively large increments in strength for a longer period than would the initially strong systems. We therefore have a basis for supposing that the speed of attainment of maximum clustering should be positively related to the degree of reinforcement prior to recall. It is necessary to add, however, that the potency of a superordinate system to induce clustering would be contingent upon a supply of subordinate systems having optimal strength. As recall proceeds, not only does this supply become progressively exhausted, but the strengths of the subordinate systems become relatively small. This would account for the final drop in clustering which occurs in our data in spite of the strength attained by the superordinate systems from our highest degree of reinforcement.

Our discussion has involved the simplification of considering as superordinate structures only those corresponding to the four categories of the words in our stimulus-list. Theoretically we have ample reason to suppose the functioning of other superordinates than these, but no reason to suppose that their existence would alter the basic character of our analysis. This theoretical account has been both tentative and sketchy. It poses additional problems for study. We believe, however, that we have demonstrated the necessity of postulating a type of mediating process to account for the phenomenon of clustering, and further, that this mediating process is modified by learning occurring during both memorization and recall. In doing this we have obtained some information on how organization operates in higher mental processes.

E. SUMMARY

The purpose of this study was to investigate the effects of reinforcement on the incidence of clustering in the recall of items presented in random order for learning. Clustering was defined as the occurrence of sequences of items in the same category. The randomized items comprised 60 nouns including 15-each of the following categories: *animals, names, professions,* and *vegetables.* Reinforcement was controlled on the basis of presenting the items from one to five times. A total of 243 undergraduate students, divided into five groups, served as subjects. The results of this study were interpreted as supporting the following experimental hypotheses:

1. The total amount of clustering occurring in the recall of randomized verbal associates should be a positive function of the degree of their reinforcement.

2. The progressive changes in clustering from the beginning to the end of recall of a group of randomized associates should be modified by reinforcement in the following ways:

a. The initial level of clustering should be a positive function of reinforcement.

b. There should be a positive relationship between reinforcement and the speed of attainment of maximum clustering.

c. For all levels of reinforcement clustering should tend to decrease as the supply of available associates approaches exhaustion.

An attempt is made to explain the phenomenon of clustering and the derivation of the experimental hypotheses in terms of Hebb's account of the development of superordinate perceptions.

REFERENCES

1. BOUSFIELD, W. A. The occurrence of clustering in the recall of randomly arranged associates. *J. Gen. Psychol.,* (in press).
2. HEBB, D. O. The Organization of Behavior. New York: Wiley, 1949. Pp. 335.
3. THORNDIKE, E. L., & LORGE, I. The Teacher's Word Book of 30,000 Words. New York: Teachers College, Columbia University, 1944. Pp. 274.

Department of Psychology
University of Connecticut
Storrs, Connecticut

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 83-109.

CORRELATIONAL ANALYSIS OF THE LEARNING AND RELEARNING OF FOUR DIFFERENT TASKS ON THE MODIFIED MASHBURN APPARATUS*

Department of Psychology, State University of Iowa

DON LEWIS, DOROTHY E. MCALLISTER, AND HAROLD P. BECHTOLDT[1]

A. INTRODUCTION

A recent investigation (2) of individual susceptibility to interference in the performance of the standard and reversed tasks on the Mashburn Apparatus, Two-Hand Coördinator, and Turret Pursuit Apparatus, yielded negative results, that is, did not reveal any consistency from device to device in the amount of interference displayed by individual subjects (*S*s). A total of 109 male students attained fairly high levels of proficiency on the standard tasks on the three devices, then received practice on the corresponding reversed tasks, and finally relearned the standard tasks. Practice on the devices was consecutive, not concurrent. Proactive interference was thought to be revealed by differences between performance levels on the first trials of standard and reversed practice and retroactive interference by losses in proficiency from the final trials of original learning to the first trials of relearning.

Several measures of interference were derived for each subject (*S*) from his performance scores on the three devices. Correlations among these measures provided no support for the hypothesis that individual male *S*s have particular degrees of susceptibility to interference which will be manifested in their performance of several different complex motor tasks.

---

*Received in the Editorial Office on January 29, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The study herein summarized was a joint venture of the authors and several research assistants. It was planned and conducted under the direction of the senior author, who undertook the preparation of the final report. The responsibility for scheduling subjects, training individual experimenters, and processing the data to the University Statistical Service fell to Dr. McAllister. Dr. Bechtoldt did the factorial analysis. The apparatus was calibrated and serviced by Mr. John P. Dolch. The subjects were run and the data posted and checked by Mr. Abram M. Barch, Mr. H. B. Dalrymple, and Mr. E. R. Dusek.

This research was supported in part by the United States Air Force under Contract No. AF 33(038)-13214, monitored by the Director of Research, Perceptual and Motor Skills Research Laboratory, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government.

Even though the proactive and retroactive changes in proficiency were large on the average as well as extensive in range, the true correlations among the measures of interference for individual *S*s could not be assumed to be other than zero.

The three possible explanations which were offered to account for the failure to obtain confirmation of the hypothesis that susceptibility to interference is a relatively stable tendency which varies in degree from person to person and which generalizes to different conditions of performance, were as follows:

1. The differences among the 109 *S*s in losses of proficiency on a task subsequent to practice on a corresponding reversed task reflected something other than real differences in susceptibility to interference.
2. Degree of susceptibility to interference was a factor in the performance of each *S* but was quite specific to each task and thus varied significantly from task to task.
3. The measures of amount of interference which entered into the calculation of the correlation coefficients were not sufficiently reliable to permit a satisfactory test of the hypothesis.

The first and third of these possible explanations seemed worthy of further study. The results of a special correlational analysis of some Mashburn data were available (2, pp.72-76). It had been predicted, on the assumption that degree of susceptibility to interference in Mashburn performance is, by and large, independent of level of attainment, that the correlation between number of matches on the last trial of *OL* and number of matches on either the first trial of *IL* or the first trial of *RL* would be relatively low. The prediction was clearly supported by the results. For example, the correlations between numbers of matches on the last two trials of *OL* and also on the first two trials of *IL* and of *RL*, were all greater than .77. In contrast, the correlations between numbers of matches on the final trial of *OL* and the first trials of *IL* and *RL* were .45 and .55, respectively. This occurred when 25 trials on the standard task and then 25 on the reversed task preceded the relearning of the standard task. In another analysis for which the number of *OL* trials was 50, the correlations between scores on the last two *OL* trials, between scores on the first two *IL* trials, and between scores on the final *OL* and first *IL* trials were .79, .81, and .30, respectively.

The sharply reduced values of the correlation coefficients over the shifts from one Mashburn task to the other were taken as added evidence of the functioning of interference at the outset of both *IL* and *RL*. This view was supported by the fact that the correlations between scores on the final *OL* trial and scores on *IL* and *RL* trials other than the first ones became increasingly large as *IL* and *RL* practice continued, and by the end of *IL* and *RL,* the intercorrelations among final trial scores were essentially as high as those between scores on any two adjacent trials in any of the three phases of learning. This trend of the correlational pattern had also been predicted.

A somewhat paradoxical finding was that the intercorrelations among scores on the first trials of the three phases of practice were all around .55. The fact that scores on the first *OL* trial were only moderately correlated with scores on either the first trial of *IL* or the first trial of *RL* was not surprising, but a correlation as low as .55 between scores on the initial trials of *IL* and *RL* needed careful examination.

Proactive interference was thought to be functioning at its greatest strength during the first *IL* trial and retroactive interference at its greatest strength during the first *RL* trial. In view of the relatively low correlation coefficient, it was necessary to infer either that proactive and retroactive interference are different phenomena, or that performances at the outset of *IL* and *RL* were influenced to an appreciable degree by unknown and perhaps irrelevant factors. If both proactive and retroactive interference consist of inappropriate response tendencies (and this seems likely), then the first alternative is unlikely, while the second alternative is at least plausible. The implication is that the changes in proficiency which had been assumed to be measures of amount of interference were not measures of interference alone but of other factors as well.

B. Plan of the Study and Predictions

A search was begun for experimental and statistical methodologies which might, in the first place, reveal something about the nature of the changes in performance which had been found in previous studies of interference and, in the second place, lead to more reliable indices of interference as such. It was conjectured that, at the point of shifting from the standard task on the Mashburn to the reversed task and also from the reversed task back to the standard task, different *S*s may respond in many different ways, each way more or less unique to a particular *S*. The relatively high correlations between scores on the first and second trials of *IL* and between scores on the first and second trials of *RL* implied that whatever the *S*s did immediately after the shifts in tasks, they did quite consistently. However, the total pattern of an individual's performance could depend upon some amount of interference and also upon varying amounts of such contaminating factors as annoyance, surprise, bewilderment, etc.

The chief aim was to find a procedure for getting "purified" measures of interference. It was felt that successive stages of practice on several different tasks on the same device might serve to eliminate most of the contaminating factors and leave interference as the principal if not sole factor determining losses in proficiency at the points of shift from one task to the next. The major apprehension was that *S*s, as they had more and more experience in shifting from task to task, might either cease completely to undergo significant losses in proficiency at the points of shifting or display differential losses which depended not so much upon original susceptibility to interference as upon differing amounts of inhibition of interfering response tendencies. However, the whole matter seemed worthy of careful study.

The decision was to provide for the learning, in succession, to fairly high levels of proficiency, of four tasks on the Modified Mashburn Apparatus (1), and then for the relearning of these same tasks. The original learning trials on all tasks except the first one constituted, in a sense, three successive phases of interpolated learning. Extended practice on the first task (Task *A*, the standard Mashburn task) was preceded by eight trials of alternating practice on this task and Task *B* [the so-called reversed Mashburn task—the task that had been used for interpolated learning in several previous studies (1, 3, 5, 6)]. The inclusion of the preliminary alternating trials was for a practical purpose. It was hoped that subsequent individual losses in proficiency at the points of shifting from one task to the next would be correlated with certain aspects of performance on the alternating trials and, consequently, that performance on a few alternating trials on the Mashburn apparatus would be indicative of individual susceptibility to interference. Spieth and Lewis (7) had previously found large and relatively stable individual differences in performance during alternating practice on the standard and reversed tasks on the Turret Pursuit Apparatus, and it seemed possible that these differences could be attributed to the interference factor.

Practice by each *S* occurred during 13 periods, each about 50 minutes in length, on 13 different days. The days were consecutive except for weekend breaks between Days 5 and 6 and Days 10 and 11. The sequence of practice on the several tasks and the number of *OL* and *RL* trials on each are shown in Table 1. There were 40 *OL* and 20 *RL* trials per task. It should be noted that the points of shifting from one task to the next always occurred at the mid-point, and never at the beginning, of a day's practice. This arrangement served to minimize if not eliminate the possibility of warm-up effects at the shifts from task to task.

TABLE 1

SEQUENCE OF PRACTICE ON THE VARIOUS TASKS

Day	Task	Phase of practice	Number of trials On indicated day	In indicated phase
1	A-B-A-B-etc.	Prelim.	8	8
"	A	OL	10	40
2	A	"	20	
3	A	"	10	
"	A–1	OL	10	40
4	A–1	"	20	
5	A–1	"	10	
"	A–2	OL	10	40
6	A–2	"	20	
7	A–2	"	10	
"	A–3	OL	10	40
8	A–3	"	20	
9	A–3	"	10	
"	A	RL	10	20
10	A	"	10	
"	A–1	RL	10	20
11	A–1	"	10	
"	A–2	RL	10	20
12	A–2	"	10	
"	A–3	RL	10	23
13	A–3	"	10	
"	A–3	"	3	

All trials were two minutes in length. On all days except Day 1, 15-second rest pauses occurred after each trial, except that a 2-minute rest was provided after each block of five trials. On Day 13, a rest period of 20 minutes followed Trial 10 on Task A-3. At the end of this rest period, three additional *RL* trials were given on the task. The purpose here was to obtain a rough estimate of the amount of warm-up decrement following a 20-minute rest subsequent to the attainment of a high level of proficiency on the Mashburn apparatus.

On Day 1, general instructions on how to perform the standard Mashburn task were given first. Then came a single trial on this task (Task *A*). At the end of the trial, the controls were reversed to obtain Task *B*, and the *S*s were instructed on the general nature of the new task. The time required for instructions was one minute. The *S*s then received alternating trials on Tasks *B* and

A with rest pauses of 15 seconds between trials. At the end of the preliminary alternating trials, there was a five-minute break. This break was followed by 10 trials on Task *A*.

The five tasks on which practice was given are described in Table 2 in terms of the direction of movement of each of the three controls. In the table, *S* stands for standard, *R* for reversed. As indicated, all of the controls had to be moved in the standard directions for Task *A*, and all in the reversed directions for Task *B*. Tasks A-1, A-2 and A-3 were obtained by reversing two controls at a time.

TABLE 2
DESCRIPTION OF TASKS IN TERMS OF DIRECTIONS OF MOVEMENT OF THE THREE CONTROLS: *S* REPRESENTS STANDARD DIRECTION, *R* REVERSED DIRECTION

	Controls		
Task	Aileron	Elevator	Rudder
A	S	S	S
B	R	R	R
A-1	R	S	R
A-2	S	R	R
A-3	R	R	S

The *S*s were all males and were volunteers from a course in elementary psychology. For participating, they received two points for each of the 13 periods of practice. These points were added to their total score on examinations and offered the possibility of a higher grade. The number of *S*s was 37. A larger number could not be obtained.

The following predictions were made:

1. The correlations between performances on the first trials of any two consecutive phases of practice will increase as practice continues. (For example, the correlation between performances on *RL* Trial 1 on Task A-2 and on *RL* Trial 1 on Task A-3 will be substantially higher than the correlation between performances on the first *OL* trials on these same two tasks.)

2. The correlations between performances on the final trial on one task and the first trial on the next will remain low, as practice continues, *provided* that the losses in proficiency at the shifts remain large.

Several expectations, all stemming from previous findings on the Mashburn apparatus, were as follows:

1. The correlations between performances on any two adjacent trials on the same task, and also between performances on the final *OL* and *RL* trials on the several tasks, will be high.

2. The correlations between performances on any two adjacent trials on a task will be higher than those between performances on remote trials on the same task.

3. The correlations between performances on the final trial on a task and performances on various trials on the next task will increase with the remoteness of the trials. (For example, the correlation between performances on the final trial on a task and the fifth trial on the next task will be substantially higher than the correlation between performances on the final and first trials.)

C. RESULTS

1. *Trends of the Performance Curves*

The trends of the performance curves are clearly depicted in Figure 1 where means of number of matches, means of total number of errors, and means of number of errors on each of the three controls are plotted against trials. There is a separate graph for each of the major phases of practice. The various curves may be identified by referring to the legend in the lower right-hand graph. Breaks in the curves, in the various graphs, signify the end of a day's practice, except for the final break in the *RL* curve for Task A-3.

The conspicuous features of the curves for matches are their general upward trends for each phase of practice, their markedly greater slopes in the later phases, the sharp decrements from the end of one phase to the beginning of the next, and the decrements which occurred, especially during the later phases, over the 24-hour periods of rest. The decrement over the 20-minute rest on the final day of practice, prior to the last three trials on Task A-3, was about 1.5 matches and was somewhat less than the typical loss over the longer periods.

There was an upward shift of the means on the first trials on the various tasks, as practice continued, and a similar upward shift in performances on the final trials. This indicated that there was some amount of proactive facilitation from the first task to the last.

In general, the curves in Figure 1 for the total errors display a downward trend from the beginning to the end of each major phase of practice. The chief exceptions are the slight increases from the first to the second trials on Tasks A-1 and A-2, and the increases over the 24-hour periods of rest. The latter increases are somewhat more noticeable during the *RL* than during the *OL* phases of practice.

An interesting fact revealed by the error curves for the separate controls is that the number of errors on the control which was not reversed in chang-

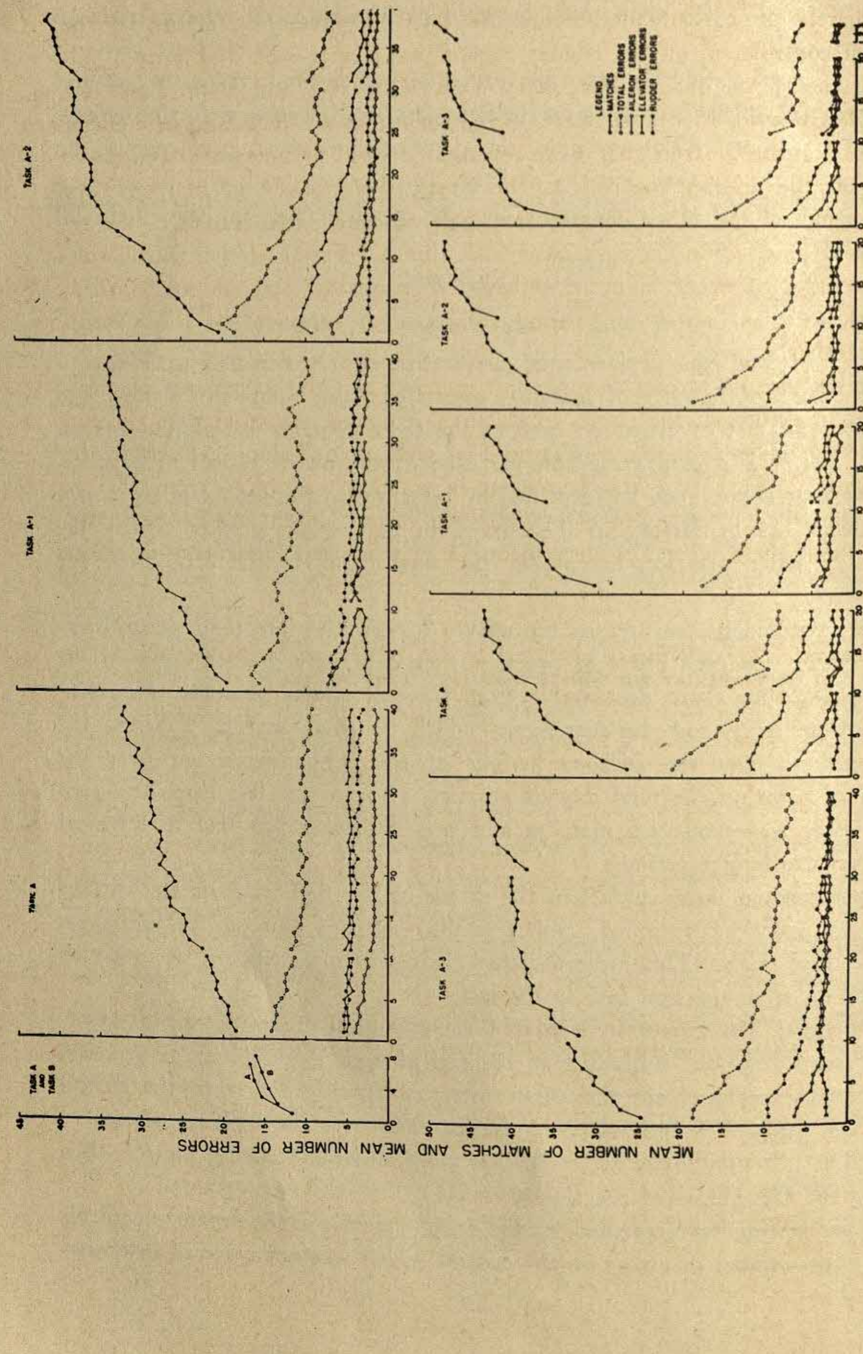


FIGURE 1

Means of Number of Matches, Means of Total Number of Errors, and Means of Errors on Each of the Three Controls Plotted Against Trials (There is a Separate Graph for Each Phase of Practice) ing from one task to the next, was either about the same or somewhat less immediately after than just before the shift. Decreases in errors on the elevator control and on the rudder control are clearly seen in Figure 1 following the shifts, during *OL,* from Task *A* to A-1 and from Task A-1 to A-2. In contrast, there were usually conspicuous increases in number of errors on the controls that were reversed. It will be noted that the errors made on the aileron control, at the shifts, were fewer in number than those made on the other control that was changed. Inadvertently, the sequence of practice on the tasks was such that a reversal of the aileron control occurred at every point of shifting. Presumably, the subjects learned, in part, to avoid errors in shifting from one arrangement of this control to the other.

The decrements in mean number of matches and the increments in mean number of errors, at the points of shifting, remained fairly large to the end of practice. This is shown in Figure 1 as well as in Table 3. Even on the final shift to Task A-3, the decrement in matches was 13.7 (from 48.3 to 34.6) and the increment in errors was 9.6 (from 6.8 to 16.4). As shown in Table 4, there were also decrements in matches and increments in errors

TABLE 3

DECREMENTS IN MEAN NUMBER OF MATCHES AND INCREMENTS IN MEAN NUMBER OF ERRORS AT THE SHIFTS FROM ONE TASK TO THE NEXT

Shift	Decrement in matches	Increment in errors
A to A-1	12.4	6.5
A-1 to A-2	13.3	8.7
A-2 to A-3	16.0	11.9
A-3 to A	16.5	13.7
A to A-1	13.0	8.9
A-1 to A-2	9.9	11.3
A-2 to A-3	13.7	9.6

TABLE 4

DECREMENTS IN MEAN NUMBER OF MATCHES AND INCREMENTS IN MEAN NUMBER OF ERRORS FROM THE FINAL *OL* TRIAL ON A TASK TO THE FIRST *RL* TRIAL ON THE SAME TASK

Task	Decrement in matches	Increment in errors
A	5.5	12.0
A-1	3.2	7.9
A-2	7.8	12.4
A-3	8.4	9.0

from the final trial in the *OL* phase of practice on a task to the initial trial in the *RL* phase on the same task. On Task A-3, for example, the retroactive decrement in matches was 8.4 (from 43.0 to 34.6), while the retroactive increment in errors was 9.0 (from 7.5 to 16.5).

The losses in proficiency at the shifts were large enough to provide for a satisfactory test of one of the major predictions—that the correlations between performances on the final and first trials on two successive tasks would remain small as practice continued.

2. *Correlational Patterns*

For studying the correlational patterns within and between the various phases of practice, Pearson product-moment coefficients were computed from the numbers of matches made by the 37 subjects on something over 100 separate trials, selected as being at critical points in the whole sequence. The error data had previously been shown to be less reliable than the data on matches and were, therefore, not utilized in the correlational analysis; but they served a useful purpose in the interpretation of the performance trends.

The intercorrelations among numbers of matches on the eight preliminary trials of alternating practice are given in the three parts of Table 5, one part for Task *A*, one for Task *B*, and a third for *A* and *B* together. The lowest coefficient is .53 (for Trials 3 and 8), the highest .82 (for Trials 4 and 8). The intercorrelations among performances on the same task are a little higher, on the average, than among those on the two different tasks. The averages are .718 and .656.

The correlations between matches on selected pairs of adjacent trials on the four tasks, during both *OL* and *RL*, are presented in Table 6. The lowest and highest values for *OL* are .67 and .95, while the lowest and highest for *RL* are .75 and .97. As will be noted, the correlations are somewhat higher for *RL* than for *OL*. However, for all phases of practice, the coefficients are usually higher than .75.

Of incidental interest, in Table 6, is the difference between the correlations for Trials 9 and 10 and those for Trials 10 and 11. Trial 10, in each phase of practice, occurred just before a 24-hour break in practice, Trial 11 just after it. The average value of the coefficients for Trials 9 and 10 is .89, for Trials 10 and 11, .83. The average difference is not large, but the separate differences are all in the same direction. The same generalization holds for differences between the coefficients for Trials 29 and 30 and those for 30 and 31. Trials 30 and 31 in *OL* were also separated by 24-hour breaks. It may reasonably be inferred from the differences in correlation coefficients that changes in set were induced during the overnight periods.

TABLE 5

INTERCORRELATIONS AMONG NUMBER OF MATCHES ON THE PRELIMINARY TRIALS ON TASK *A*, ON TASK *B*, AND ON TASKS *A* AND *B*

	Task *A*					Task *B*					Tasks *A* and *B*			
Trials	1	3	5	7	Trials	2	4	6	8	Trials	1	3	5	7
1		67	70	70	2		81	57	71	2	67	68	67	63
3			80	68	4			75	82	4	64	66	71	71
5				78	6				63	6	62	67	77	61
7					8					8	61	53	68	64

TABLE 6
CORRELATIONS BETWEEN NUMBERS OF MATCHES ON SELECTED PAIRS OF ADJACENT TRIALS ON THE FOUR TASKS DURING THE ORIGINAL LEARNING AND RELEARNING PHASES OF PRACTICE

		Pairs of trials									
	Task	1 and 2	2 and 3	3 and 4	4 and 5	9 and 10	10 and 11	19 and 20	29 and 30	30 and 31	39 and 40
Original learning	A	77	76	73	75	81	72		89	77	94
	A-1	70	87	80	67	79	73		91	81	91
	A-2	68	79	81	88	88	78		91	88	95
	A-3	69	79	78	87	86	82		93	89	94
Relearning	A	75	86	90	94	96	91	93			
	A-1	84	94	90	86	90	85	92			
	A-2	85	93	92	91	94	92	93			
	A-3	88	91	94	96	95	90	97			

The general decrease in the value of the correlation coefficient with remoteness of trials is shown in Table 7 and also in Table 8. The decrease is revealed most unequivocally, in both tables, by the values for the original learning phase on Task *A*, which occurred prior to extended practice on any of the other tasks. In Row 1 of Table 7, the first value of .77 is for *OL* Trials 1 and 2 on Task *A*. The other values in this row show a general decrease, the last value for Trials 1 and 40 being .58. This obvious decrease with remoteness of trials is made even more emphatic in Table 8 where the correlations are for numbers of matches on the final trial of a task and on several preceding trials on the same task. In Row 1, the correlations of .94 and .85 are for Trials 40 and 39 and Trials 40 and 30, respectively, on Task *A*. As in Table 7, Row 1, the value for Trials 40 and 1 is .58. The decrease with remoteness is seen to be quite large.

In every row in Tables 7 and 8, the last value is smaller than the first. However, except for Task *A* in original learning, the decreases are to be attributed not only to remoteness but also to interference and perhaps other factors. The amount of interference was presumably greatest during the first trial in each major phase of practice and least during the last trial. Thus, the change in the correlation coefficients arising from a dissipation of interference was bound to be in the same direction as any change accompanying remoteness itself.

The changes in the correlation coefficients across the rows in Table 7 are much smaller for the *RL* phases of practice than for the *OL* phases. The difference is especially marked in the case of Task A-3. As seen, the

TABLE 7

CORRELATIONS BETWEEN THE NUMBER OF MATCHES ON THE FIRST TRIAL ON EACH TASK AND NUMBER OF MATCHES ON LATER TRIALS ON THE SAME TASK*

		Trial							
	Task	2	3	5	10	15	20	30	40
Original learning	A	77	75	72	72		55	60	58
	A-1	70	72	71	66		41	51	23
	A-2	68	67	68	44		36	27	26
	A-3	69	70	67	55		35	36	32
Relearning	A	75	78	70	61	62	62		
	A-1	84	83	73	69	67	61		
	A-2	85	84	79	75	77	75		
	A-3	88	85	91	83	78	81		

*Read across the table. A given row contains the correlations between matches scores on the first trial and matches scores on later trials on the task specified in the first column.

TABLE 8

CORRELATIONS BETWEEN THE NUMBER OF MATCHES ON THE FINAL TRIAL OF A TASK AND THE NUMBER OF MATCHES ON EACH OF SEVERAL PRECEDING TRIALS ON THE SAME TASK*

		Trial								
	Task	39	30	20	19	15	10	5	3	1
Original learning	A	94	85	83			74	61	69	58
	A-1	91	84	80			54	44	22	23
	A-2	95	87	87			69	57	55	26
	A-3	94	93	94			78	73	53	32
Relearning	A				93	92	88	83	80	62
	A-1				92	90	84	76	69	61
	A-2				93	90	90	83	85	75
	A-3				97	98	94	91	90	81

*Read across the table. A given row contains the correlations between matches scores on the final trial and matches scores on preceding trials on the task specified in the first column.

correlations for *OL* Trials 1 and 2 and 1 and 20 are .69 and .35, respectively. In contrast, the values for *RL* Trials 1 and 2 and 1 and 20 are .88 and .81.

The rôle of interference in shaping the correlational patterns is further suggested by the values in Table 9. In this case, the coefficients are for numbers of matches on the final trial on one task and on selected trials on the succeeding task. For example, the several correlations in the first row are for Trial 40 on Task *A* and Trials 1, 2, 3 . . . 10, etc. on Task A-1.

TABLE 9

Correlations Between the Number of Matches on the Final Trial of an Immediately Preceding Task and the Number of Matches on Each of Several Trials on the New Task*

Preceding task	New task	Trials 1	2	3	4	5	10	20	30	40
A	A-1	31	47	39	38	52	60	72	78	78
A-1	A-2	35	56	59	66	63	80	88	83	86
A-2	A-3	34	60	53	64	67	73	90	92	91
A-3	A	61	65	81	81	81	90	91		
A	A-1	70	75	75	84	83	84	87		
A-1	A-2	74	83	84	85	86	90	82		
A-2	A-3	84	89	87	91	91	89	91		

*Read across the table. A given row contains the correlations between matches scores on the final trial on the task specified in the first column and selected trials on the task specified in the second column.

The lowest value of .31 reflects the presence of factors (including interference) in performance on the first trial on Task A-1 which were largely absent in performance on the last trial on Task *A*. As these factors became less influential with continuing trials on Task A-1, the correlations became larger and larger until, finally, the correlation between matches on the 40th trials on Tasks *A* and A-1 was .78. This value compares favorably in magnitude with some of the coefficients for pairs of adjacent trials on the same task—and so do the values of .86 and .91 (the last ones in Rows 2 and 3) for the 40th *OL* trials on Tasks A-1 and A-2 and the 40th *OL* trials on Tasks A-2 and A-3.

The trends of the correlation coefficients in the first three rows of Table 9 were predicted from previous information on Mashburn practice. The trend of the coefficients for the final trial on Task A-3 and subsequent relearning trials on Task *A* was similar in kind, but the value of .61, for Trial 40 on A-3 and Trial 1 on *A,* was much larger than anticipated. The three values of .70, .74, and .84, appearing in Table 9 immediately below .61, were wholly unexpected. They are contradictory of the most important single prediction made before the investigation was begun—that if losses in proficiency remained large at the shifts from one task to the next, the correlation between the number of matches on the final trial of the task and on the initial trial on the new task would be low. The correlation of .84 at the last shift from A-2 to A-3 is surprisingly high. Its high value suggests either that the individual losses in proficiency at the shift were not due to interference or that they were highly related to individual levels of proficiency which, in turn, were the chief determiners of amount of interference.

TABLE 10

CORRELATIONS BETWEEN NUMBER OF MATCHES ON THE LAST *OL* TRIAL ON EACH TASK AND NUMBER OF MATCHES ON EACH OF SEVERAL *RL* TRIALS ON THE SAME TASK*

	RL trial							
Task	1	2	3	4	5	10	15	20
A	55	51	60	67	65	75	78	79
A-1	55	69	72	74	77	81	80	83
A-2	70	78	86	80	86	84	88	87
A-3	86	88	90	93	93	94	93	95

*Read across the table. A given row contains the correlations between matches scores on the final *OL* trial and matches scores on selected *RL* trials on the task specified in the first column.

The values in the last four rows of Table 9 are entirely consistent with those in Table 10, where the correlations are between performances on the last *OL* trial on a task and on selected *RL* trials on the same task. The trends of the coefficients, especially in the top two rows, are similar to trends previously obtained for the relearning of Task *A* subsequent to interpolated practice on Task *B*. The only difference of importance is that the first values of .55 are somewhat higher. The values for Task A-3, in the bottom row, display a slight upward trend, but they are all so high as to preclude any inference that the factors contributing to final *OL* and subsequent *RL* proficiency on Task A-3 were appreciably different. The results for this task are clearly not what were predicted.

The change toward higher and higher correlations, as practice continued, is emphasized in Table 11 where intercorrelations are presented for the first trials and also for the final trials in the various phases of learning.

TABLE 11

INTERCORRELATIONS—ABOVE THE DIAGONAL, AMONG NUMBERS OF MATCHES ON FIRST TRIALS, AND BELOW THE DIAGONAL, AMONG NUMBERS OF MATCHES ON THE FINAL TRIALS OF THE VARIOUS PHASES OF PRACTICE

		Prelim. phase	*OL* phases				*RL* phases			
			A	A-1	A-2	A-3	A	A-1	A-2	A-3
	Prelim.		50	38	19	06	22	11	27	33
OL	A	48		51	47	30	40	47	48	48
	A-1	34	78		41	51	32	16	14	27
	A-2	28	76	86		54	68	39	47	34
	A-3	38	82	83	91		55	42	30	37
RL	A	24	79	83	91	91		70	80	68
	A-1	22	73	83	88	87	87		71	79
	A-2	22	76	73	87	89	93	82		79
	A-3	35	80	79	87	95	91	83	91	

First trial intercorrelations appear above the diagonal, last trial intercorrelations below. The intercorrelations for first trials are generally low (one as low as .06, another .11) except when both trials fall in *RL* phases of practice. The only really low intercorrelations for final trials are those involving Trial 8 of preliminary practice, the final trial on Task *B*. The remaining ones, 28 in number, are all .73 or higher, and only seven are below .80.

Additional correlations of interest are given in Table 12. Included here are coefficients for number of matches on the final trial of each phase of practice and number on the first trial of each subsequent phase. For example, performance on the final preliminary trial was correlated with performance on each of eight subsequent first trials. The coefficients are shown in the first row of the table. It was predicted at the outset of the investigation that, although the intercorrelations among performances on the final trials in the various practice phases would all be relatively high, the correlations between last and first trial performances would remain relatively low. Actually, as seen in Table 12, the last and first trial correlations became quite large, especially in the relearning phases. This finding suggests the possibility that interference became less and less important as a determiner of performance level at the points of shifting from one task to the next.

TABLE 12

CORRELATIONS BETWEEN THE NUMBER OF MATCHES ON THE FINAL TRIAL OF EACH MAJOR PHASE OF PRACTICE AND THE NUMBER OF MATCHES ON THE FIRST TRIAL OF EACH SUCCEEDING PHASE

	A	A-1	A-2	A-3	A	A-1	A-2	A-3
Prelim.	42	58	41	11	38	15	37	40
A		31	34	27	55	53	64	75
A-1			35	37	53	55	69	74
A-2				34	61	70	70	79
A-3					61	71	75	86
A						70	73	85
A-1							74	78
A-2								84

The coefficients of correlation presented in several of the previous tables show a general trend toward higher values as practice continued. This trend suggests that the higher correlations may have been dependent, at least in part, on level of proficiency. As already shown in Figure 1, the mean number of matches became higher during the later stages of practice.

To make a gross check on the possible relationship between degree of correlation and level of proficiency, intercorrelations were run among number of matches on several selected trials, regardless of their position in the whole sequence, on which the means were approximately the same and, in one case, all near the mid-level, and, in the second case, all at a relatively high level.

Five trials from five different phases of practice were selected for the first group of intercorrelations. They are identified and their means given in Table 13, where the intercorrelations are also presented. The means fell around 29. The intercorrelations range from .50 to .75.

TABLE 13
INTERCORRELATIONS AMONG NUMBER OF MATCHES ON SELECTED TRIALS ON WHICH THE MEANS WERE ALL NEAR MID-LEVEL

	Task	Phase of practice	Trial	Mean number of matches	1	2	3	4	5
1	A	OL	30	28.8		72	60	58	47
2	A-1	OL	20	30.0			75	69	50
3	A-2	OL	10	29.8				64	63
4	A-3	OL	4	28.8					64
5	A	RL	2	29.3					

Six trials located in six different phases of practice, the identities and means of which appear in Table 14, were selected to obtain intercorrelations among performances at relatively high levels. In this case, the means of matches were at 40 or a little above. The intercorrelations range from .82 to .92, with an average value of .863. This average is substantially higher than the one of .625 for the intercorrelations in Table 13.

TABLE 14
INTERCORRELATIONS AMONG NUMBER OF MATCHES ON SELECTED TRIALS ON WHICH THE MEANS WERE ALL RELATIVELY HIGH

	Task	Phase of practice	Trial	Mean number of matches	1	2	3	4	5	6
1	A-2	OL	39	40.5		89	89	84	85	88
2	A-3	OL	30	40.1			92	87	86	84
3	A	RL	15	42.1				85	82	86
4	A-1	RL	10	40.0					86	85
5	A-2	RL	5	40.3						86
6	A-3	RL	3	40.7						

The question arises as to why the correlations were generally higher among performances at the higher levels of attainment. A possible explanation was that the standard deviation of number of matches per trial increased with increases in the means of matches. It is well known that, other things equal, a correlation coefficient is likely to be higher if there is an extension of the range of the measures being correlated. This is especially true if the number of cases is not large.

TABLE 15
MEANS AND STANDARD DEVIATIONS OF NUMBER OF MATCHES ON REPRESENTATIVE TRIALS IN THE VARIOUS PHASES OF PRACTICE

Task	Phase of practice	Trial	Mean	*SD*
A	Prelim.	1	11.6	3.10
B		2	13.4	3.05
B		6	15.2	2.60
A		7	16.8	2.60
B		8	16.0	2.76
A	OL	1	18.5	3.42
A		2	19.1	3.57
A		3	19.4	3.60
A		5	20.3	2.97
A		10	22.0	3.41
A		20	26.6	4.55
+A		30	28.8	4.55
A		40	32.0	4.96
A-1	OL	1	19.6	2.68
A-1		2	20.9	2.74
A-1		3	21.5	3.20
A-1		5	22.6	3.06
A-1		10	25.1	3.64
+A-1		20	2[illegible].9	4.04
A-1		40	33.7	5.08
A-2	OL	1	20.4	3.61
A-2		2	22.6	3.58
A-2		3	23.8	3.57
A-2		5	25.3	4.20
+A-2		10	29.8	4.93
A-2		20	35.8	6.10
++A-2		39	40.5	7.38
A-2		40	40.7	6.35

Task	Phase of practice	Trial	Mean	*SD*
A-3	OL	1	24.7	3.54
A-3		2	27.1	3.50
A-3		3	27.9	3.95
+A-3		4	28.8	4.53
A-3		5	30.3	4.13
A-3		10	33.3	5.46
A-3		20	38.9	6.81
++A-3		30	40.1	6.92
A-3		40	43.0	7.65
A	RL	1	26.5	4.30
+A		2	29.4	4.65
A		3	31.1	4.82
A		5	33.1	5.67
A		10	38.2	7.53
++A		15	42.2	7.46
A		20	43.5	6.97
A-1	RL	1	30.5	4.37
A-1		2	34.0	5.45
A-1		3	35.4	5.43
A-1		5	36.7	5.85
++A-1		10	40.0	5.94
A-1		20	42.8	7.17
A-2	RL	1	32.9	4.71
A-2		2	37.0	6.30
A-2		3	38.5	6.43
++A-2		5	40.3	7.04
A-2		10	44.0	7.90
A-2		20	48.3	8.70
A-3	RL	1	34.6	5.36
A-3		2	38.9	6.26
++A-3		3	40.7	6.84
A-3		5	41.9	7.91
A-3		10	44.4	8.99
A-3		20	48.5	9.77

+Trial involved in intercorrelations given in Table 13.
++Trial involved in intercorrelations given in Table 14.

Information concerning the magnitudes of the standard deviations on representative trials in the various phases of practice is given in Table 15, along with the means on the trials. It is seen that the standard deviations do have a marked tendency to be large whenever the means are large. Interestingly enough, the average value of the *SD*'s for the five trials involved in the intercorrelations presented in Table 13 is 4.54 while the average for the six trials involved in Table 14 is 6.92. The difference between the average values of the intercorrelations in the two tables might have depended to an appreciable degree on the difference in the *SD*'s. Unfortunately, there is no satisfactory way of testing this possibility. However, it is certain that the correlational patterns previously discussed cannot be attributed solely to changes in *SD*'s. Much evidence in support of this assertion may be found by studying the coefficients in Tables 6 and 9, along with the *SD*'s in Table 15. To give but a single example: The correlation between matches on *OL* Trial 40 on Task *A* and *OL* Trial 1 on Task A-1 is .31. The *SD*'s for the trials are 4.96 and 2.68. In contrast, the correlation between matches on *OL* Trials 1 and 2 on Task A-1 is .70. The *SD*'s in this case are 2.68 and 2.74. Many similar examples may be found.

Perhaps the best position to take, at this point, is that the magnitudes of the obtained correlation coefficients were probably affected to some extent by variations in the *SD*'s, but that the underlying correlational patterns were determined principally by differences in the factorial structure of performance at various stages of practice. It is possible, of course, that this position is a reasonable one only for the *OL* phases of practice and becomes increasingly untenable from the first to the last phase of relearning. A final interpretation must take account of the results of a factorial analysis.

3. *Factorial Analysis*

As a first step in securing information on the possible factors contributing to performance at various points in practice, intercorrelations among numbers of matches on 20 selected trials were obtained. Four of these trials were the 1st, 6th, 7th, and 8th of preliminary practice, while the other 16 were the first and last trials on each of the four *OL* phases and each of the *RL* phases of practice. The matrix of intercorrelations is given, above the diagonal, in Table 16.

The matrix was first factored to four factors by Thurstone's multiple group method (8). An additional centroid factor was then determined from the 4th factor residuals. A second cycle of the multiple group method using all five factors was then carried out to stabilize the communalities.

TABLE 16

Intercorrelations Among Numbers of Matches on Twenty Selected Trials, Fifth Factor Residuals, and Loadings on Five Factors

	Task	Phase of practice	Trial																					Loadings				
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	A	B	C	D	E
1	A	Prelim.	1		62	70	71	50	46	38	37	19	28	06	36	22	32	11	16	27	29	33	40	78				
2	B	"	6	04		61	63	51	28	56	20	37	11	26	16	30	15	21	01	21	08	27	15	69		—12	38	
3	A	"	7	06	—05		64	72	39	57	35	49	23	25	28	33	24	17	13	27	19	31	29	69			36	30
4	B	"	8	—03	03	—02		42	48	58	34	41	28	11	38	38	24	15	22	37	22	40	35	67	24			
5	A	OL	1	02	—01	10	—08		58	51	46	47	37	30	45	40	41	47	36	48	41	48	45	44		27	41	30
6	"	"	40	—03	—01	—05	02	06		31	78	34	76	27	82	55	79	53	73	64	76	75	80	25		68		21
7	A-1	"	1	—04	01	—07	14	—05	00		23	41	17	51	23	32	18	16	07	14	07	27	17	48			56	35
8	"	"	40	01	03	—01	—02	—04	—02	—05		35	86	37	83	53	83	55	83	69	73	74	79			76		36
9	A-2	"	1	00	—02	02	03	—01	03	—07	—01		26	54	23	68	18	39	20	47	26	34	19		47		50	40
10	"	"	40	00	02	01	01	—05	—04	02	02	00		34	91	61	91	70	88	70	87	79	87			86		16
11	A-3	"	1	01	03	—05	—02	—10	—03	12	00	05	01		32	55	37	42	25	30	32	37	27			26	54	35
12	"	"	40	—02	—03	—01	02	—02	—01	03	—01	—02	00	01		61	91	71	87	75	89	86	95	11		85		
13	A	RL	1	02	—02	01	01	—06	00	01	—04	02	00	06	—01		62	70	57	80	65	68	56		46	37	33	
14	"	"	20	01	01	01	—03	—04	—01	01	01	—04	—01	04	—03	01		70	87	73	93	85	91		—11	90		
15	A-1	"	1	—01	00	00	—02	08	—01	—05	—01	00	03	—02	02	—04	—03		61	71	75	79	67			60	35	
16	"	"	20	—05	01	—02	00	02	—03	00	01	—04	00	—04	00	00	00	00		74	82	78	83	—13		84	—14	19
17	A-2	"	1	—02	—01	01	—03	08	—02	—03	02	00	—03	—04	—01	01	00	02	02		75	79	70		40	52		
18	"	"	20	04	—03	03	—01	04	03	—04	—03	04	00	00	—01	00	01	—01	—01	00		84	91			84		
19	A-3	"	1	—03	00	00	02	00	01	02	00	—02	—02	00	00	—04	—01	04	02	01	—01		81	12		70	13	
20	"	"	20	00	—03	02	00	02	00	00	—02	02	—01	01	02	00	—02	01	—02	—02	03	—02		14		84		

Intercorrelations

Residuals

The 5th factor residuals, shown below the diagonal in Table 16, ranged from +.14 to —.10, with most of them (177 out of 190) falling between ±.05.

The five factors were rotated by the method of radial rotation and two dimensional plots to a "simple structure" configuration, as defined by Thurstone (8). The resulting configuration was oblique with the correlations between the primary vectors ranging from —.34 to +.43. Four of the planes were clearly determined with at least 9 of the 20 loadings between ±.10. One plane had only five values within this "zero" range and must be regarded as poorly defined. The evaluation of the configuration as a whole, using both the criteria of zero loadings and of positive manifold, indicates that a "meaningful structure" (in the sense in which Thurstone uses this term) was obtained. Only four negative values greater than —.10 appeared in the solution. The largest of these was —.14.

The loadings for Factor *A* are largest for the first seven trials listed at the left in Table 16. The first five of the seven consisted of practice on Tasks *A* and *B*, in alternation. The factor is identified as *initial adjustment* to the task situation.

Factor *C* has the highest loadings on the final trials in the various phases of practice. The loadings on the first trials become higher as practice continues but within each phase of practice, the loading is markedly higher on the last than on the first trial. This factor is identified as *coördination of responses* at relatively high levels of proficiency. It is the factor largely responsible for determining individual performance asymptotes.

Factor *D* is identified as *susceptibility to interference*. Except for the loadings on Trials 6 and 7 of preliminary practice, all other loadings of consequence for Factor *D* are on first trials in the various phases. As practice continues, the loadings become negligible. This fact is taken to mean that interference, as a phenomenon independent of coordination of responses, decreased until it was of little importance. In other words, Factor *D* was not an independent detectable source of variance in performance on *RL* Trial 1 on Tasks A-2 and A-3. The losses in proficiency which occurred at the final two points of shift in task (see Table 3 and the curves in Figure 1) were apparently due in very large part to a factor which produced the variability of the trial means about the grand mean of all trials within a phase of practice. This variability was not a part of the variability on any single trial and, therefore, was not reflected in the correlation coefficients in the matrix in Table 16.

Factor *B* is somewhat anomalous. However, the three highest loadings are on the first *OL* trial on Task A-2, the first *RL* trial on this same task, and the first *RL* trial on Task *A*. In all three of these shifts in task, the aileron and elevator controls were both reversed as compared with their setting on the preceding task. In all other shifts in task beyond the preliminary phase of practice, the reversals were on the aileron and rudder controls. Factor *B* is presumably specific to a double reversal on the stick. It is tentatively identified as *difficulty in making this particular type of adjustment.*

Factor *E* cannot be identified. The loadings are probably residuals.

Factor *D*, the interference factor, fails to show a loading on Trial 8 of preliminary practice. This fact, when considered along with the loadings of .38 and .36 on preliminary Trials 6 and 7, is interpreted as indicating that performance on the preliminary trials cannot be utilized in predicting individual susceptibility to interference, that is, in predicting performance at the points of shifting from one task to another. At least, multiple regression procedures would not be any more fruitful for prediction than the zero-order correlations between performances on the preliminary trials and performances on the various first trials following shifts in task.

D. DISCUSSION

The general trends of the correlational data were as expected, with one important exception. It was predicted that if the losses in proficiency (as represented by decreases in matches and/or increases in errors) remained large at the shifts from task to task in the later phases of practice, the correlations between final trials on one task and the initial trials on the next would be low. Actually, these correlations became increasingly high, as seen in Table 9. The prediction was based on the assumptions (*a*) that the losses would be directly proportional to amounts of interference and (*b*) that interference would be a relatively independent source of variance—independent, for example, of final level of performance.

Another prediction—that the correlations between matches scores on the first trials of two consecutive phases of practice would become increasingly high as practice continued—was borne out by the results. However, the basis for the prediction may have been an erroneous one. It was assumed that as practice on the various tasks continued, the losses in proficiency at the shifts would come to reflect interference *only* and not interference plus other incidental and irrelevant factors. The losses at the last two shifts (from Task A-1 to A-2 and from A-2 to A-3, on Days 11 and 12) may have depended upon interference, but they must, at the least, be regarded as covariant with levels of proficiency.

The foregoing discussion is supported by the results of the factorial analysis (refer to Table 16). Three of the five factors were apparently of major importance in determining performance level at different points in the practice sequence. They are identified as (*A*) *initial adjustment* to the task situation; (*C*) *coördination of responses* at relatively high levels of proficiency; and (*D*) *susceptibility to interference.* The contribution of Factor *A* presumably became negligible by the end of the *OL* phase on Task A-1. The interference factor (*D*) has substantial loadings on the initial trials of the four *OL* phases of practice and the first two *RL* phases, but the loadings are inconsequential for the last two *RL* phases. Factor *C* has maximal loadings on the final trials of the eight major phases of practice, but the loadings for the factor on the first trials of the *RL* phases are clearly on the upswing. It is obvious that during the *RL* phases of practice on Tasks A-2 and A-3, Factor *C* played a predominant rôle while Factor *D* played either a negligible or a dependent rôle. There is no empirical basis, aside from the losses in proficiency at the shifts, for assuming the functioning of the interference factor during the final phases of *RL* practice.

The objection might be raised that the number of subjects ($N = 37$) was not sufficiently large to provide unequivocal information on the correlation trends or the underlying factors in performance. It is true, of course, that single correlation coefficients based on an N of 37 are not very dependable. However, the trends of the present data leave little doubt concerning the major conclusions. A locus of possible uncertainty lies in the upward change in the magnitude of the variances from the early to the late stages of practice. The effects of large changes in variance on correlation coefficients are not readily determinable. It is possible that the upward change in overall variance (which must be attributed mainly to Factor *C*) was so great as to obscure the variance of the factor identified as interference. It seems far more plausible to believe, however, that interference virtually disappeared as a factor independent of Factor *C*. This does not mean that interference (consisting of tendencies to make inappropriate responses) no longer was present. Rather, it means that amount of interference for individual subjects was highly correlated with degree of coördination of responses. Losses in proficiency at the last two shifts in tasks have, therefore, no independent predictive value.

The conclusions reached in the present investigation are supported, in large part, by the results of a companion study, soon to be reported (4), which involved several successive phases of practice by 43 *S*s on the standard and reversed tasks on the *SAM* Complex Coördinator. The correlational patterns are fundamentally alike, and the same three factors—initial adjustment, interference, and coördination of responses—were identified as being the primary determiners of performance level at comparable points in the whole sequence of trials. Unfortunately, in both studies the variances of number of matches per trial increased significantly from the early to the late stages of practice, so the uncertainty stemming from the change in variances cannot be empirically dispelled without further experimentation.

A principal aim of the investigation was to discover a "purified" measure of interference. The expectation was that successive shifts from task to task would serve to eliminate extraneous and irrelevant sources of variance (would, in effect, reduce individual variability in work methods, confusion, annoyance, etc.) and leave amount of interference as the major determiner of performance level. The empirical evidence of this expected change was to be a significant increase, as practice continued, in the correlation between matches scores on the first trials following any two successive shifts in task. The increase occurred only as between the *OL* and *RL* phases of practice. This is shown in Table 11. The correlations during *OL* were .51, .41, and .54. During *RL,* they were .70, .71, and .79. The difference cannot be ascribed to a less contaminated functioning of the interference factor but must be attributed to the increasing predominance of Factor *C*.

E. SUMMARY

Thirty-seven male *S*s were given extended practice on four different tasks on the Modified Mashburn Apparatus. The tasks included standard Task *A* and three others (A-1, A-2, and A-3) obtained by reversing two controls at a time. There was an *OL* phase of 40 trials and an *RL* phase of 20 trials on each task. These major phases of practice were preceded by a preliminary phase consisting of eight trials during which each *S* practiced alternately on Task *A* and Task *B* (the complete reversal of *A*).

The trials, two minutes in length, were separated by rest pauses of 15 seconds, except that a rest of two minutes followed each block of five trials. Practice occurred during 13 sessions lasting about 50 minutes each, on 13 days which were consecutive except for weekends. No more than 20 trials were given in any one session. Each shift in task occurred at the mid-point of a practice session.

The measures of performance were number of matches and number of errors per trial—total number of errors as well as the number on each of the three controls. The means of matches and errors were plotted to show graphically the trends of performance. The matches scores on representative trials throughout the performance sequence were used in computing correlation coefficients. The coefficients served to reveal trends in correlational patterns, and some of them entered into a factorial analysis of performance.

The aims of the experiment were to observe the effects of successive shifts in task on the amount of interference, to seek a "purified" measure of interference, and to discover the other fundamental factors contributing to performance level. The preliminary phase of practice was included to see if scores on a few alternating trials on the standard and reversed Mashburn tasks could be used to predict amounts of interference later displayed by individual *S*s.

The principal results may be summarized as follows:

1. The means of number of matches increased and the means of total errors decreased in a fairly regular way during each major phase of practice. The chief deviations from regularity occurred immediately after the overnight breaks in practice. On the first trial of each day's session (except during *OL* on the first task), there was a decrease in matches and an increase in errors. These losses in proficiency, which were identified as warmup decrements, were relatively small.

2. Substantial losses in proficiency occurred at the shifts from one task to the next. Even at the final point of shifting on Day 12, the loss was large.

3. Despite the losses in proficiency at the successive shifts, the overall trend of performance was upward. This meant that while the absolute losses at the shifts remained about the same, the relative losses became smaller.

4. The increases in errors at the shifts were due to increased numbers of errors on the two controls that were reversed. The number of errors on the control that was not reversed did not increase.

5. The correlations between matches scores on representative pairs of adjacent trials in the major phases of practice were generally high. Sixty-four such correlations were computed. Of these, only three were less than .70 and only 17 were less than .80, while 28 were .90 or higher. The reliability of the individual matches scores was deemed reasonably satisfactory.

6. Within each phase of practice, there was a decrease in the correlation between performances on pairs of trials as the trials being correlated became increasingly remote. This decrease was particularly pronounced within the *OL* phases on Tasks A-1, A-2, and A-3, at the outset of which interference was presumably at maximal strength. The decreases were smaller for the *RL* phases of practice.

7. The 28 intercorrelations among the numbers of matches on the final trials of the eight major phases of practice were generally high. None of the 28 was lower than .73 and only seven were below .80. This indicated that the four tasks were basically alike, in a performance sense.

8. During the *OL* phases of practice, as fully expected, the correlations between numbers of matches on the final trial of a task and the first trial of the next task were low. The correlations increased more or less regularly as scores on trials farther and farther beyond the first were correlated with scores on the preceding final trial.

9. During the *RL* phases of the practice and particularly during the last *RL* phase, the correlations between matches scores on the final trial of a task and the first trial of the next were *not* low. This was contrary to expectations and occurred despite the substantial losses in proficiency at the points of shifting.

10. As predicted, the correlations between matches scores on the first trials of two consecutive phases of practice were substantially higher toward the end than at the beginning of the practice sequence. In view of the results of the factorial analysis, this change did not necessarily signify that irrelevant factors became inconsequential as practice continued.

11. Factorial analysis yielded five factors, one of which was apparently residual. The other four factors were identified. Three of these were particularly important in determining performance level at various points in the practice sequence. Factor *A*—initial adjustment to the task situation—contributed maximally in the preliminary phase and the first *OL* phase. Factor *D*—susceptibility to interference—played a highly significant rôle on the initial trials of the *OL* phases and the first two *RL* phases of practice but apparently ceased to contribute in any independent way during the last two *RL* phases. Factor *C*—coördination of responses at high levels of proficiency—contributed maximally, at first, to performances on the final trials on the various tasks but toward the end of the practice sequence, it became the predominant, if not the sole, factor determining level of proficiency.

12. The scores on the eight trials of alternating practice in the preliminary phase provided no reliable basis for predicting individual losses in proficiency at the points of shifting—the points at which interference was assumed to be maximal.

13. The attempt to discover a "purified" measure of interference was not successful. There were no significant loadings for the interference factor during the last phases of practice and the only evidence that interference was present at all consisted of the losses in proficiency at the shifts in task. These losses, even though large, were too closely related to Factor *C* to be regarded as suitable measures of amount of interference.

REFERENCES

1. LEWIS, D., & SHEPHARD, A. H. Devices for studying interference in psychomotor performance: I. The modified Mashburn apparatus. *J. of Psychol.*, 1950, **29**, 35-46.
2. LEWIS, D., & MCALLISTER, D. E. An investigation of individual susceptibility to interference. U. S. Navy, Special Devices Center, *Technical Report* SDC 938-1-10, 1950.
3. LEWIS, D., MCALLISTER, D. E., & ADAMS, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. Exp. Psychol.*, 1951, **41**, 247-260.
4. LEWIS, D., MCALLISTER, D. E., & BECHTOLDT, H. P. Correlational study of performance during successive phases of practice on the standard and reversed tasks on the *SAM* complex coördinator. *J. of Psychol.*, 1953, **36**, 111-126.
5. MCALLISTER, D. E., & LEWIS, D. Facilitation and interference in performance on the modified Mashburn apparatus: II. The effects of varying the amount of interpolated learning. *J. Exp. Psychol.*, 1951, **41**, 356-363.
6. SHEPHARD, A. H. Losses of skill in performing the standard Mashburn task arising from different levels of learning on the reversed task. U. S. Navy, Special Devices Center, *Technical Report* SDC 938-1-9, 1950.
7. SPIETH, W., & LEWIS, D. The effects of alternating practice on the performance of two antagonistic motor tasks. U. S. Navy, Special Devices Center, *Technical Report* SDC 938-1-6, 1950.
8. THURSTONE, L. L. Multiple-Factor Analysis. Chicago: Univ. Chicago Press, 1947.

Department of Psychology
State University of Iowa
Iowa City, Iowa

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 111-126.

CORRELATIONAL STUDY OF PERFORMANCE DURING SUCCESSIVE PHASES OF PRACTICE ON THE STANDARD AND REVERSED TASKS ON THE *SAM* COMPLEX COORDINATOR*

Department of Psychology, State University of Iowa

Don Lewis, Dorothy E. McAllister, and Harold P. Bechtoldt[1]

A. Summary of Companion Study

In a companion study of the present one (2), 37 male subjects were given extended practice on four different tasks on the Modified Mashburn Apparatus. After a preliminary phase of practice consisting of eight alternating trials on the standard and reversed tasks, each subject received in succession 40 *OL* (original learning) trials and 20 *RL* (relearning) trials on each of the principal tasks. Relatively high levels of proficiency were attained on each task prior to the shift to the next one.

The purpose of the study was to examine the basic features of performance through the application of correlational procedures and factorial analysis. Of particular interest were the changes in the correlational patterns and the accompanying changes in the factorial composition of performance. Several predictions were made before the study began. Two predictions which were of crucial importance were (*a*) that the correlation between performance scores on the first trials of two successive phases of practice would be substantially higher toward the end than at the beginning of the whole practice sequence, and (*b*) that if the losses in proficiency at the shifts in task remained large as practice continued, the correlation between performance scores on the final trial of one task and the first trial of the next would be low.

The first prediction was based on the assumptions that early in the practice sequence, level of performance on the first trial on a task would depend not only on an interference factor but also on incidental or irrelevant factors such as annoyance, confusion, surprise, etc.; that factors other than interference would become increasingly less important as practice continued, and that the principal, if not the sole, factor determining level of performance on first trials toward the end of the whole sequence would be interference. The second prediction was based on the empirical fact that correlations between performance scores on the final and first trials at the points of shifting from one task to the next on the Modified Mashburn Apparatus are relatively low when only *OL, IL,* and *RL* phases of practice are provided, and on the assumptions that losses in proficiency at the shifts after extended practice on several different tasks would signify the presence of interference, and that the interference factor would be the chief determiner of level of performance on a first trial but would be inconsequential in determining level on an immediately preceding final trial.

---

*Received in the Editorial Office on January 29, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Assistance in making the study was given by Abram M. Barch, John P. Dolch, E. R. Dusek, and Howard H. McFann.

This work was supported in part by the United States Air Force under Contract No. AF 33(038)-13214, monitored by the Director of Research, Perceptual and Motor Skills Research Laboratory, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government.

The first prediction turned out to be correct; the correlations between first trial performances on successive phases of practice increased substantially as practice continued. However, the change apparently occurred for reasons other than those originally inferred. The factorial analysis revealed that the interference factor virtually ceased to be an independent contributor to performance prior to the last two *RL* phases and that level of performance, toward the end of the practice sequence, was determined chiefly by a coördination factor. This held for first trials as well as final trials on a task. The predominance of a single factor in the later phases of *RL* was fully anticipated, prior to the completion of factorial analysis, when it was seen that the correlations between performance scores on final and first trials were quite high. This, of course, was contrary to the second prediction. Even though the losses in proficiency were substantial at the two final shifts in task, the analysis did not support the view that interference was an independent source of variance.

B. Design of Present Study

The present study involved successive phases of *OL* and *RL* on only two tasks—the standard and reversed tasks—on the *SAM* Complex Coördinator. The purpose was basically the same as that of the companion study on the Mashburn: to analyze performance trends and discover the underlying factors in performance. In the Mashburn study, there were four successive phases of *OL* (or, from one point of view, three phases of interpolated learning following a single phase of *OL*) prior to four successive phases of *RL*. In the present study, there were successive phases of learning and relearning on two different tasks (or several phases of *RL* on two tasks subsequent to the *OL, IL,* and *RL* phases of the common retroactive inhibition paradigm).

The two predictions of greatest importance were exactly the same as the two given above for the Mashburn study. Actually, the predictions for the two studies were made at the same time; the two were planned concurrently.

Table 1 shows the sequence of practice on the two tasks. *A* and *B* stand for standard and reversed tasks, respectively. There was a preliminary phase consisting of eight alternating trials on the two tasks. These trials were included to see if scores on them could be used to predict the extent of individual losses at the points of shifting after extended practice on one or the other of the tasks. There were two *OL* phases of practice of 40 trials each and then three *RL* phases on each task, all but the last of which consisted of 20 trials. Practice occurred during 10 sessions on 10 different days, consecutive except for a week-end break of two days. The shifts in task always occurred at the midpoints of practice sessions, never at the be-

TABLE 1
SEQUENCE OF PRACTICE ON THE TWO TASKS

Day	Task	Phase of practice	Number of trials	
			On indicated day	In indicated phase
1	A-B-A-B, etc.	Prelim.	8	8
“	A	OL	10	40
2	A	“	20	
3	A	“	10	
“	B	OL	10	40
4	B	“	20	
5	B	“	10	
“	A	RL-1	10	20
6	A	“	10	
“	B	RL-1	10	20
7	B	“	10	
“	A	RL-2	10	20
8	A	“	10	
“	B	RL-2	10	20
9	B	“	10	
“	A	RL-3	10	20
10	A	“	10	
“	B	RL-3	10	10

ginning. This arrangement was used to prevent any confusion that might have arisen if warmup decrements had occurred along with losses due to interference. The trials were all two minutes in length and were separated by 15-second rest pauses except for a two-minute rest after each block of five trials.

The total number of *S*s was 43. They were volunteers from an introductory course in psychology but received bonus points, which had a bearing on their final grade, for each period they served.

The data used in the analysis consisted of the number of matches made on each trial by each *S*. The means for trials were utilized in preparing graphs to depict general trends in performance. The scores of individual *S*s on representative trials entered into the calculation of Pearson product-moment correlation coefficients. Some of the coefficients served to reveal correlational patterns and some were used in making a factorial analysis.

C. RESULTS

1. *Performance Curves*

In Figure 1, the means of number of matches are plotted against trials. There are separate graphs for the various phases of practice. In fact, to obtain a more symmetrical figure and also to emphasize losses at the shifts in task, there are two parts of the graph for the first *RL* phase on Task *A*. The breaks in the performance curves within the various graphs represent day-to-day breaks in practice. It should be noted that the ordinates of all of the lower graphs are shortened.

As seen in the figure, the trends of the performance curves are generally upward during each practice phase. The only exceptions of importance are the losses which appear on the first trials after the 24-hour breaks. These losses were presumably related to a need for warming up. The *S*s kept on improving despite the shifts in task; the means on the first trials and also on the final trials of the various phases became progressively higher as practice continued.

The losses in proficiency that occurred at the shifts in tasks remained large to the end of the practice sequence. The amounts of loss are given in Table 2. The loss of 16.8 matches at the final shift was comparable in magnitude to those occurring at the four immediately preceding shifts, and was large enough to provide for a satisfactory test of the second prediction—that the correlation between scores on the final and first trials at the shift would be low.

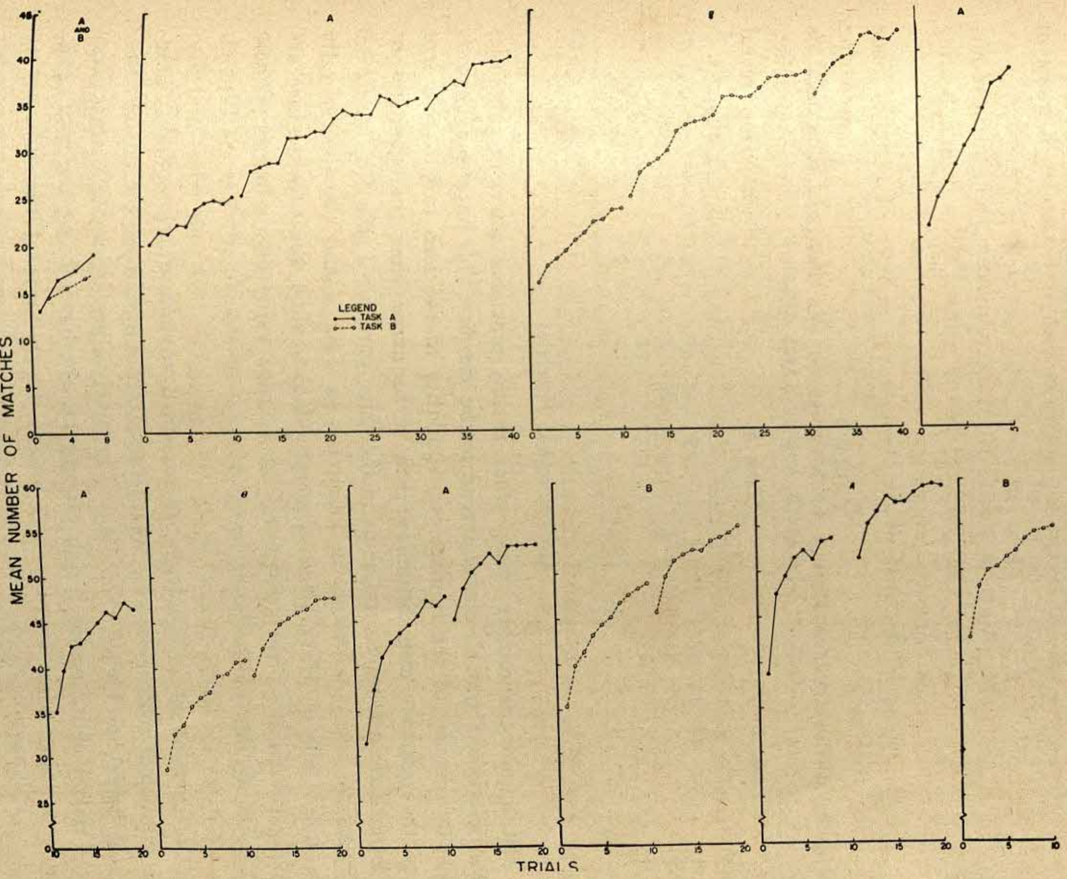


FIGURE 1
Means of Number of Matches Plotted Against Trials in Each of the Nine Phases of Practice

TABLE 2

DECREMENTS IN MEAN NUMBER OF MATCHES (AT THE SHIFTS) FROM THE FINAL TRIAL ON ONE TASK TO THE FIRST TRIAL ON THE NEXT

Shift	Decrement
A (OL) to B (OL)	24.1
B (OL) to A (RL-1)	21.0
A (RL-1) to B (RL-1)	17.9
B (RL-1) to A (RL-2)	16.1
A (RL-2) to B (RL-2)	18.0
B (RL-2) to A (RL-3)	16.4
A (RL-3) to B (RL-3)	16.8

TABLE 3

INTERCORRELATIONS AMONG NUMBER OF MATCHES ON THE PRELIMINARY TRIALS ON TASK *A*, ON TASK *B*, AND ON TASKS *A* AND *B*

Task *A*					Task *B*					Tasks *A* and *B*				
Trials	1	3	5	7	Trials	2	4	6	8	Trials	1	3	5	7
1		63	63	56	2		80	66	67	2	55	62	62	62
3			79	81	4			76	76	4	51	65	72	66
5				85	6				75	6	39	59	64	57
7					8					8	39	65	69	64

2. *Correlational Patterns*

During the preliminary phase, when the *S*s practiced alternately on the two tasks, there was an upward trend in the means. This improvement was expected because of the fundamental similarity of the two tasks. The fact that the tasks are also dissimilar, in a performance sense, may be readily inferred from the intercorrelations in the three parts of Table 3. The intercorrelations among scores on the four trials on Task *A* and also those among the four trials on Task *B* are generally higher than the intercorrelations among the trials on the two tasks, the average value of the former being .723, the average of the latter being .594.

The correlations in Table 4 are presented to indicate the reliability of the individual matches scores. With five exceptions, all of the correlations in the table are higher than .80. The five values below .80 are either for Trials 1 and 2 or for Trials 10 and 11, the latter being the trials just before and just after an overnight break. Thirty-three of the 56 correlations are higher than .89. The values for the final two trials in the several phases of practice are all .91 or higher. The reliability of the individual scores is deemed adequate.

The strengths of the relationships among performances on the first trials of each phase of practice and selected later trials in the same phase are shown

TABLE 4

CORRELATIONS BETWEEN NUMBERS OF MATCHES ON SELECTED PAIRS OF ADJACENT TRIALS IN VARIOUS PHASES OF PRACTICE

		Pairs of trials									
Task	Phase of practice	1 and 2	2 and 3	3 and 4	4 and 5	9 and 10	10 and 11	19 and 20	29 and 30	30 and 31	39 and 40
A	OL	83	83	85	89	91	86		89	88	93
B	OL	81	87			88	76		93	86	91
A	RL-1	86	87	83	89	94	74	95			
B	RL-1	74	86	90	94	92	79	91			
A	RL-2	78	93	90	94	95	90	95			
B	RL-2	91	96	93	94	95	91	94			
A	RL-3	85	92	93	95	94	90	92			
B	RL-3	94	87	92	93	96					

TABLE 5

CORRELATIONS WITHIN EACH PHASE OF PRACTICE BETWEEN THE NUMBER OF MATCHES ON THE FIRST TRIAL ON A TASK AND NUMBER OF MATCHES ON LATER TRIALS ON THE SAME TASK

		Trial							
Task	Phase of practice	2	3	5	10	15	20	30	40
A	OL	83	86	82	80	79	72	74	73
B	OL	81	81	76	60	64	46	39	19
A	RL-1	86	78	60	36	21	25		
B	RL-1	74	80				59		
A	RL-2	78	78	70			51		
B	RL-2	91	90	85	82	69	65		
A	RL-3	85	78	76	74	74	81		
B	RL-3	94	91	94	91				

by the coefficients in Table 5. The trend toward lower correlations with remoteness of the trials correlated, even when the interference factor is presumed to be relatively unimportant, is shown by the values in the first row. The decrease was very large (from .81 to .19) during *OL* practice on Task *B,* when interference was presumably of maximal importance in determining level of performance on Trial 1. It will be observed that the amount of change with remoteness decreased as practice continued from one *RL* phase to the next. As seen, there was virtually no change over the 10 trials in the third *RL* phase on Task *B.*

The correlational trend within each phase of practice may also be seen in Table 6. The coefficients in this case were obtained by correlating scores on the final trial of a task with scores on several preceding trials on the same task. The decrease with the remoteness of the trials correlated that occurred before interference became an important factor was from .93 to

TABLE 6

CORRELATIONS WITHIN EACH PHASE OF PRACTICE BETWEEN THE NUMBER OF MATCHES ON THE FINAL TRIAL ON A TASK AND THE NUMBER OF MATCHES ON EACH OF SEVERAL PRECEDING TRIALS ON THE SAME TASK

Task	Phase of practice	Trials 39	30	20	19	15	10	5	3	1
A	OL	93	89	88		80	71	68	73	73
B	OL	91	84	83		68	54	33	25	19
A	RL-1				95	93	75	52	29	25
B	RL-1				91	92	84	73	63	59
A	RL-2				95	93	92	86	75	51
B	RL-2				94	96	87	82	79	65
A	RL-3				92	91	91	90	87	81
B	RL-3							94	93	91

TABLE 7

CORRELATIONS BETWEEN THE NUMBER OF MATCHES ON THE FINAL TRIAL IN THE IMMEDIATELY PRECEDING PHASE OF PRACTICE AND THE NUMBER OF MATCHES ON EACH OF SEVERAL TRIALS IN THE NEW PHASE

Preceding phase and task		New phase and task		1	2	3	5	10	15	20	30	40
OL	A	OL	B	16	22	22	31	51		72	75	85
OL	B	RL-1	A	26	30	32	45	72	86	87		
RL-1	A	RL-1	B	53	49	57	71	80	80	85		
RL-1	B	RL-2	A	58	68	74	81	81	88	85		
RL-2	A	RL-2	B	70	80	83	89	90	87	86		
RL-2	B	RL-3	A	78	84	86	92	88	89	90		
RL-3	A	RL-3	B	87	90	91	89	90				

.73, as shown in the first row of the table. As seen in the second and third rows, the drops were large (from .91 to .19 and from .95 to .25) after interference became a crucial factor in determining levels of performance subsequent to a shift. It is apparent, once again, that the amount of change with remoteness became less as practice continued beyond the first *RL* phase on Task *A*.

A check on the correctness of the second prediction is readily made with the coefficients of correlation in Table 7. These coefficients were obtained by correlating the scores on the final trial of one task with scores on representative trials in the next task. The prediction was that the correlation between performances on the final and first trials at the final point of shifting would be low if the loss in proficiency was large. As already indicated, the average loss for the 43 *S*s was 16.8 matches. The correlation over the shift was .87. Thus, the prediction was clearly incorrect. The correlations between performances over the shifts became increasingly large as practice proceeded. The low values of .16 and .26 at the first two shifts confirm previous findings on the Mashburn apparatus (1) as well as the findings of the companion study (2). It was assumed, when the second prediction was made, that the correlations at later points of shifting would be of the same general order of magnitude as those at earlier points if the losses in proficiency were equally large.

TABLE 8

INTERCORRELATIONS: ABOVE THE DIAGONAL, AMONG NUMBERS OF MATCHES ON THE FIRST TRIALS, AND BELOW THE DIAGONAL, AMONG NUMBERS OF MATCHES ON THE FINAL TRIALS OF THE VARIOUS PHASES OF PRACTICE

	Prelim.	A (OL)	B (OL)	A (RL-1)	B (RL-1)	A (RL-2)	B (RL-2)	A (RL-3)	B (RL-3)
Prelim.		49	—01	03	36	39	49	50	43
A (OL)	62		23	26	46	59	66	53	62
B (OL)	59	85		63	51	48	35	39	26
A (RL-1)	58	88	87		52	53	38	35	27
B (RL-1)	63	76	83	85		69	77	68	62
A (RL-2)	56	83	80	90	85		78	78	68
B (RL-2)	61	78	83	86	93	86		82	76
A (RL-3)	56	75	77	83	83	86	90		82
B (RL-3)	53	73	79	81	90	86	91	90	

The correctness of the first prediction is confirmed by the intercorrelations in Table 8. Values for the first trials in the various phases of practice are given above the diagonal while values for final trials appear below the diagonal. Correlations between first trials on successive phases appear just above and along the diagonal and range from .23 for the *OL* phases on the two tasks to .88 for the *RL*-3 phases. The predicted increase occurred. The notion was that the increase would depend upon a gradual elimination of irrelevant or incidental factors in performance and on a corresponding increase in the relative importance of the interference factor. This intepretation was apparently erroneous.

A clear indication that the interference factor became less and less influential as practice continued is given by the values in Table 9. In this case, the correlations are between performances on the last trials of the various phases of practice and the first trials of all succeeding phases. Except for the values in Row 1, which bear on performance on the last trial of the preliminary phase of practice, the values in each row are much larger for the later phases. In the second row, for example, the change is from .16 to .80; in the third row, from .26 to .80; in the fifth, from .58 to .84, and so on. This signifies that performance on the final and first trials became more

TABLE 9

CORRELATIONS BETWEEN THE NUMBER OF MATCHES ON THE FINAL TRIAL OF EACH PHASE OF PRACTICE AND THE NUMBER OF MATCHES ON THE FIRST TRIAL OF EACH SUCCEEDING PHASE

	Prelim.	A (OL)	B (OL)	A (RL-1)	B (RL-1)	A (RL-2)	B (RL-2)	A (RL-3)	B (RL-3)
Prelim.		66	48	36	59	63	60	64	59
A (OL)			16	26	50	52	72	64	80
B (OL)				26	52	59	73	71	80
A (RL-1)					53	52	70	75	80
B (RL-1)						58	71	78	84
A (RL-2)							70	73	82
B (RL-2)								78	88
A (RL-3)									87
A (RL-3)									

closely related as practice continued and, consequently, that interference became less influential in determining performance on first trials.

A fact of some importance is that the correlations are all relatively high for pairs of trials on which the *S*s displayed relatively high levels of proficiency. It seemed appropriate, therefore, to examine the variability of individual scores on representative trials to see if changes in variability were occurring concomitantly with changes in performance level.

Table 10 presents the means and standard deviations of number of matches on representative trials in the various phases of practice. It is obvious at once that there is a positive relationship between the means and *SD*'s. The effects of changes in standard deviations on correlation coefficients are not easy to assess. There is the possibility that the higher correlations for trials on which proficiency was great were due in part to the larger *SD*'s. But it is unreasonable to assume that changes in *SD*'s accounted for the systematic changes in correlational patterns. Many of the correlations in Table 4 between scores on adjacent trials are quite high even though the standard deviations of the scores are low. For example, the correlation for Trials 1 and 2 in the first *RL* phase on Task *A* is .86. The two *SD*'s are 4.21 and 4.25. In contrast to this, the correlation between the matches scores on the final trial in the *OL* phase on Task *B* and the first trial in Phase *RL*-1 on Task *A* is only .26. The two *SD*'s in this case are 7.34 and 4.21. Additional examples of the same sort can be found by interrelating values in Tables 4, 7, and 10.

The evidence suggests that changes in the variability of the scores may have had a bearing on the size of the correlation coefficients but that the chief determiner of size was the degree of similarity in the factorial com-

TABLE 10
MEANS AND STANDARD DEVIATIONS OF NUMBER OF MATCHES ON REPRESENTATIVE TRIALS IN THE VARIOUS PHASES OF PRACTICE

Task	Phase of practice	Trial	Mean	*SD*
A	Prelim.	1	13.2	3.50
B	Prelim.	2	14.7	3.06
B	Prelim.	6	16.7	3.16
A	Prelim.	7	19.1	3.85
B	Prelim.	8	17.7	2.80
A	OL	1	20.1	3.98
A	OL	2	21.2	4.25
A	OL	3	21.2	4.07
A	OL	5	21.9	4.38
A	OL	10	25.1	4.46
A	OL	20	31.9	5.69
A	OL	30	35.4	5.73
A	OL	40	39.9	7.85
B	OL	1	15.8	3.46
B	OL	2	17.6	3.05
B	OL	3	18.3	3.11
B	OL	5	20.2	3.50
B	OL	10	23.6	3.97
B	OL	20	33.2	6.40
B	OL	30	37.9	6.16
B	OL	40	42.2	7.34
A	RL-1	1	21.2	4.21
A	RL-1	2	24.2	4.25
A	RL-1	3	25.9	4.42
A	RL-1	5	29.7	5.84
A	RL-1	10	38.0	7.14
A	RL-1	20	46.6	7.80
B	RL-1	1	28.7	4.96
B	RL-1	2	32.7	6.15
B	RL-1	3	33.8	6.62
B	RL-1	5	36.8	7.10
B	RL-1	10	40.9	8.29
B	RL-1	20	47.5	8.79
A	RL-2	1	31.4	6.38
A	RL-2	2	37.5	8.41
A	RL-2	3	41.0	8.45
A	RL-2	5	43.8	9.19
A	RL-2	10	47.8	9.68
A	RL-2	20	53.3	10.31
B	RL-2	1	35.3	8.92
B	RL-2	2	39.9	10.02
B	RL-2	3	41.3	9.56
B	RL-2	5	44.3	9.71
B	RL-2	10	48.9	10.76
B	RL-2	20	55.1	11.44
A	RL-3	1	38.7	8.60
A	RL-3	2	47.5	10.95
A	RL-3	3	49.5	10.77
A	RL-3	5	52.3	11.50
A	RL-3	10	54.2	11.33
A	RL-3	20	59.4	12.20
B	RL-3	1	42.6	9.90
B	RL-3	2	48.2	10.18
B	RL-3	3	50.0	10.48
B	RL-3	5	51.5	11.57
B	RL-3	10	54.9	12.71

position of performance at various points in the practice sequence. This view is supported, of course, by the results of the factorial analysis.

3. *Factorial Analysis*

Intercorrelations among the matches scores on 20 selected trials were used in making a factorial analysis of performance. Four of the trials were the 1st, 6th, 7th, and 8th of the preliminary phase of practice. The other 16 were the first and final trials in the two *OL* and the six *RL* phases of practice. The matrix of intercorrelations appears in Table 11, above the diagonal.

The matrix was analyzed, in two cycles, to three factors by Thurstone's multiple group method (3). The third factor residuals are entered in

TABLE 11

INTERCORRELATIONS AMONG NUMBERS OF MATCHES ON TWENTY SELECTED TRIALS WITH THIRD FACTOR RESIDUALS AND LOADINGS ON THREE FACTORS

Task	Phase of practice	Trial	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Loadings A	C	D
1 A	Prelim.	1		39	56	39	49	49	—01	43	03	47	36	45	39	50	49	47	50	45	43	43	22	17	
2 B	"	6	01		57	75	55	54	28	59	23	57	42	59	45	51	47	65	56	55	53	51	37		
3 A	"	7	10	—05		64	83	72	31	61	29	64	37	55	55	62	58	61	58	58	61	62	61		
4 B	"	8	—02	16	—08		66	62	48	59	36	58	59	63	63	56	60	61	64	56	59	53	45		26
5 A	OL	1	05	—04	08	—02		73	23	58	26	64	46	61	59	65	66	62	53	58	62	60	51		
6 "	"	40	—01	—07	—02	—04	02		16	85	26	88	50	76	52	83	72	78	64	75	80	73	33	31	
7 B	OL	1	—13	01	05	05	—04	—01		19	63	17	51	17	48	05	35	14	39	21	26	18			70
8 "	"	40	—04	02	—04	—02	—06	07	00		26	87	52	83	59	80	73	83	71	77	80	79	15	46	
9 A	RL-1	1	—11	—05	02	—05	—02	05	11	03		25	52	21	53	20	38	18	35	26	27	19			60
10 "	"	20	—02	00	—02	—03	—02	07	03	06	05		53	85	52	90	70	86	75	83	80	81	15	50	
11 B	RL-1	1	06	—02	—05	03	02	02	—02	—02	01	00		59	69	52	77	55	68	59	62	54		39	59
12 "	"	20	—02	04	—05	05	00	—02	00	02	—01	00	02		58	85	71	93	78	83	84	90		61	
13 A	RL-2	1	05	—04	04	00	07	—02	—07	02	00	—04	—03	—01		51	78	54	78	55	68	58		30	57
14 "	"	20	01	—04	00	00	03	03	—04	—01	04	05	02	—01	—01		70	86	73	86	82	86		57	
15 B	RL-2	1	08	—07	01	—04	08	06	—07	04	—05	00	07	—01	06	02		65	82	66	76	66		44	38
16 "	"	20	—03	09	—03	03	—02	—04	03	00	01	—02	02	05	—01	—03	—06		78	90	88	91		59	
17 A	RL-3	1	07	02	02	01	—05	—06	00	—03	—06	—01	—04	—01	05	—02	03	01		81	82	80		55	35
18 "	"	20	—01	02	01	00	—01	—01	04	—03	03	—01	01	—03	—04	01	—06	03	02		87	90		64	
19 B	RL-3	1	—03	—02	02	00	02	04	02	00	—02	—03	—01	—01	04	—01	00	02	00	02		91		61	13
20 "	"	10	—04	—02	05	—01	02	—05	06	—02	00	—05	—02	02	02	—01	—06	01	01	02	05			67	

Intercorrelations

Residuals

Table 11 below the diagonal. They range from —.13 to +.16, with all but 13 of them falling between —.06 and +.07.

The method of radial rotation and two-dimensional plots was employed in obtaining what Thurstone calls a "simple structure" configuration. The obtained configuration was oblique. The correlations between the primary vectors were .19, .31, and .75. The three planes were fairly well determined with at least six of the 20 loadings lying between ±.10. No negative loading greater than —.10 appeared in the solution. It was felt, at the end of the analysis, that a meaningful structure had been obtained.

The loadings for the three factors appear at the extreme right in Table 11. Loadings falling within the range ±.10 are not shown. Most of the loadings of consequence for Factor *A* are on the four trials in the preliminary phase and the two trials in the first *OL* phase. The factor is identified as *initial adjustment* to the task situation.

With one exception, the significant loadings for Factor *D* are on the first trials of the major phases of practice. The factor is immediately identified as *susceptibility to interference.* The loadings became progressively smaller as practice continued with .13 as the value for the first trial in the final *RL* phase on Task *B*. The change in loadings for the factor is taken to indicate that interference virtually ceased to be an independent determiner of performance level by the end of the practice sequence.

The loadings for Factor *C* are highest on the final trials of the various phases of practice. After the first RL phase on Task *A,* there are also loadings for the factor on the first trials of the successive phases. Toward the end of practice, the loadings on first and final trials are approximately of equal magnitude. The factor is identified as *coördination of responses* at relatively high levels of proficiency. It is the factor believed to be chiefly responsible for determining individual differences in performance asymptotes.

Factor D, the interference factor, had no loadings greater than .10 on the 1st, 6th, and 7th trials of the preliminary phase of practice and a loading of only .26 on the 8th trial. It is concluded, therefore, that scores on the trials in the preliminary phase cannot be used to predict individual losses in proficiency at the points of shifting from one task to another.

The three factors obtained in the present analysis are the same as the three principal ones (*A, C,* and *D*) obtained in the analysis of the companion study (2). The structure is somewhat simpler in the present case, but the results of the two analyses are fundamentally alike.

D. Discussion

Only a very brief discussion is needed here in view of the rather lengthy one in the companion paper (2). The first prediction—that the correlation between scores on the first trials of two successive phases of practice would be substantially higher toward the end than at the beginning of the practice sequence—was correct. The reason for this change, however, was erroneously identified. The expectation was that irrelevant or incidental factors would initially play some part in shaping performances on first trials but would gradually become negligible as interference became of predominant importance. The results indicate that interference itself become of greatly decreased consequence and that another factor—coördination of responses at relatively high levels of performance—became the predominant, if not sole, determiner of performance level.

The second prediction—that late in the performance sequence, the correlations between scores on final and first trials at the shifts in task would be low if the losses in proficiency were large—did not turn out to be correct. The correlations were actually fairly large and reflected the predominance of the coördination factor not only on final trials but on first trials as well. Despite substantial losses at the shifts, the interference factor apparently did not function independently to create them. At best, it co-varied with the coördination factor.

The performance curves, the correlational patterns, and the results of the factorial analysis of the present study were basically like those of the companion study. This was not entirely unexpected even though the four tasks on the Mashburn apparatus were obtained by reversing two controls at a time while the two tasks on the Complex Coördination Test were completely reversed forms of each other. All tasks on both devices are obviously very similar. The reason for running both studies was that on the Mashburn, four tasks were available to provide four phases of *OL* prior to *RL* and both matches and error scores could be obtained, while on the Complex Coördination Test only successive *OL* and *RL* phases on two tasks could be provided and only matches scores could be secured. It was felt that the two sets of results would complement one another, as they do.

E. Summary and Conclusions

A group of 43 male *S*s received extended practice on the standard and reversed tasks on the *SAM* Complex Coördination Test. There was an *OL* phase of 40 trials on each task followed by three *RL* phases on each. All of the *RL* phases except the final one on the reversed task consisted of 20 trials. The eight major phases of practice were preceded by a preliminary phase in which the *S*s practiced alternately for a total of eight trials on the two tasks.

The trials were two minutes in length and were separated by pauses of 15 seconds except that two-minute rests followed blocks of five trials. Practice was given during 10 sessions of about 50 minutes each, on 10 days which were consecutive except for one week-end break of two days. The maximum number of trials during any one session was 20. Shifts in task occurred at the mid-points of practice sessions.

The measures of performance were number of matches per trial. The means of matches were plotted to obtain performance curves. Individual matches scores on trials selected from the whole sequence were used in computing Pearson *r*'s. The *r*'s in various combinations revealed correlational patterns, and some were used in making a factorial analysis.

The aims of the study were to observe the changes in amount of interference with successive shifts in task and to discover the other fundamental factors that determined level of performance. The eight trials in the preliminary phase were included to see if scores on them could be used to predict individual losses in proficiency that occurred later at the shifts in task.

The principal results were as follows:

1. The means of the matches scores increased quite regularly during each major phase of practice. The chief exception to regular increases occurred on the first trials after the 24-hour breaks in practice. Losses on these trials were identified as warm-up decrements. They were relatively small.

2. Large losses in proficiency occurred at the shifts in task. Even on Day 10 at the final point of shifting, the loss was substantial.

3. Despite large losses at the shifts, the trend of the performance curves was upward. Initial and final means in one phase of practice were higher than initial and final means in preceding phases.

4. The correlations between matches scores on representative pairs of adjacent trials were generally high. Thus, the reliability of the individual scores was deemed adequate.

5. Within each practice phase, the correlations between scores on pairs of trials decreased with the remoteness of the trials correlated. The decrease occurred in the first *OL* phase prior to the functioning of a sizeable amount of interference. The decrease was large once interference had been generated. It became progressively smaller after the first *RL* phase.

6. The intercorrelations among matches scores on the final trials of the major phases of practice were all .73 or higher. This indicated that individual performances on the two tasks at high levels were quite similar.

7. At the first two shifts in task, the correlations between scores on the final and first trials were low. The correlations increased as scores on trials farther and farther beyond the first were correlated with scores on the preceding final trial. This finding was fully expected. At the last three points of shift in task, the correlation between scores on final and first trials were *not* low. This was contrary to one of the more important predictions.

8. The correlations between matches scores on the first trials of two successive phases of practice were much higher toward the end than at the beginning of the practice sequence. This change was predicted but it apparently arose for reasons other than the one originally specified.

9. The factorial analysis yielded three factors: (*A*) initial adjustment to the task situation; (*C*) coördination of responses at relatively high levels of performance, and (*D*) susceptibility to interference. Toward the end of the practice sequence, Factor *D* ceased to play an independent rôle while Factor *C* became the predominant determiner of level of performance.

10. The scores on the eight preliminary trials could not be used to make reliable predictions concerning amount of interference displayed by individual *S*s at later shifts in task.

REFERENCES

1. LEWIS, D., & MCALLISTER, D. E. An investigation of individual susceptibility to interference. U. S. Navy, Special Devices Center, *Technical Report*, SDC 938-1-10, 1950.
2. LEWIS, D., MCALLISTER, D. E., & BECHTOLDT, H. P. Correlational analysis of the learning and relearning of four different tasks on the Modified Mashburn Apparatus. *J. of Psychol.*, 1953, **36**, 83-109.
3. THURSTONE, L. L. Multiple-Factor Analysis. Chicago: Univ. Chicago Press, 1947.

Department of Psychology
State University of Iowa
Iowa City, Iowa

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 127-130.

A NON-PARAMETRIC TEST FOR THE RELIABILITY OF THE DIFFERENCE BETWEEN TWO PROPORTIONS*

Department of Psychology, University of California

J. A. GENGERELLI AND JOHN L. MICHAEL

Although it has been urged at least once (1), the statistical textbooks ignore the quadratic method for determining the confidence limits of a proportion. The essence of the method is to solve for p in the equation

$$(1) \qquad \frac{p_0 - p}{\left[\frac{p(1-p)}{N}\right]^{1/2}} = k$$

where p_0 is the obtained proportion and k is the ratio or coefficient representing a desired confidence level. The solution is given by

$$(2) \qquad p = \frac{(2Np_0 + k^2) \pm [(2Np_0 + k^2)^2 - 4(N + k^2)(Np_0^2)]^{1/2}}{2(N + k^2)}$$

The solution given by the quadratic method differs but little from that yielded by the usual procedure as long as p_0 is not far from .5. However, as p_0 moves away from .5, the discrepancy between the confidence limits yielded by the two methods increases. For example when $p_0 = .10$, with $N = 100$ and $k = 2$, the quadratic method yields the limits .054 and .176 while the usual formula gives .04 and .16. The difference even here is not dramatic; however, it is clear that the limits given by the conventional method are in error, since for example, the limit .176 yields $k = 2$ exactly, whereas the limit .160 does not, but rather the value $k = 1.64$. Similarly, the limit .054 gives $k = 2$, exactly whereas the limit .04 gives $k = 3.06$.

The greatest advantage of the quadratic method is to be found in connection with the reliability of the difference between two empirical proportions. This aspect of the matter is not treated by Lindquist.

In calculating the reliability of the difference between two proportions, the circularity involved in the conventional method is particularly undesirable, since we make an estimate of the parametric proportion under the null hypothesis on the basis of a weighted mean of the two obtained pro-

---

*Received in the Editorial Office on February 4, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

portions. It is clearly more desirable to be able to answer the question: *is there a parametric proportion which could, at a given confidence level, yield the obtained difference between two empirical proportions* than the question: *could the obtained difference have come from a universe having as its parametric proportion a value equal to the weighted mean of the empirical proportions?*

Let there be two empirical proportions p_1 and p_2, based, respectively, on N_1 and N_2 cases. Let ρ be the correlation between the two proportions and let the confidence coefficient be k.

Then we will have, under the null hypothesis,

(3)
$$\frac{p_1 - p_2}{\left[\frac{pq}{N_1} + \frac{pq}{N_2} - 2\rho\left(\frac{pq}{N_1}\right)^{1/2}\left(\frac{pq}{N_2}\right)^{1/2}\right]^{1/2}} = k$$

$$\frac{p_1 - p_2}{\left[(p - p^2)\left(\frac{N_1 + N_2}{N_1 N_2} - \frac{2\rho}{N_1^{1/2} N_2^{1/2}}\right)\right]^{1/2}} = k$$

where p is the parametric proportion sought under the null hypothesis.

For simplicity of algebraic manipulation, set $p_1 - p_2 = \delta$ and set the expression in the second parenthesis in the denominator equal to α. Then we have

(4)
$$\frac{\delta}{[(p - p^2)\alpha]^{1/2}} = k.$$

Squaring, transposing, and collecting,

(5)
$$k^2\alpha p^2 - k^2\alpha p + \delta^2 = 0.$$

If we solve (5) for p in the usual manner, we have,

(6)
$$p = \frac{k^2\alpha \pm [k^4\alpha^2 - (4)(k^2\alpha)(\delta^2)]^{1/2}}{2k^2\alpha}$$

It will be noted that (6) is, with one exception, expressed solely in terms of known quantities. It contains the difference between the two empirical proportions, N_1 N_2, k and ρ, the correlation coefficient between the two proportions. This latter is the only dubious quantity, since, of course, its value is unknown. In the case of independent proportions, $\rho = 0$ and hence the second term in the second parenthesis in the denominator of equation (3) disappears. In this case, all the values in (6) are known.

For given values of N_1 and N_2 and for a given difference between p_1 and p_2 (i.e., for a given δ) the solution of (6) gives unvarying results for any fixed confidence level. Thus, if k, N_1 and N_2 be fixed, pairs of obtained proportions such as .05-.15; .15-.25; .25-.35; .35-.45, etc., will be equally significant or insignificant, as the case may be. This, of course, is not the case when the usual method is used, since here the "estimated p" would change in value from one pair to the other, causing the estimated variance of the sampling distribution to increase as we move toward $pq = .25$. Hence the significance of the differences .05-.15 would be overestimated and that of .35-.45, underestimated.

The solution for p by equation (6) yields real or imaginary numbers. If the two values are real, they represent the limits of a range of proportions which, at the specified confidence level, could generate the obtained difference. In this instance, the null hypothesis would be sustained. As an example, let $p_1 = .55$, $p_2 = .45$, $N_1 = N_2 = 100$, $k = 2$, and $\rho = 0$. In this instance,[1] $p = .146$ and $.853$. Thus at the confidence level proposed, any parent population in which the proportion is greater than .146 and less than .853 could generate the obtained difference. Or, differently expressed, any single parametric proportion from .146 to .853 would yield a critical ratio of 2 or less.

On the other hand, let $p_1 = .60$ and $p_2 = .40$, with the other values remaining the same. Substitution in (6) now gives an imaginary number. This is to be interpreted to mean that, at the confidence level indicated, there is no population which could generate the obtained difference. Or, in different language, there is no proportion which will yield a critical ratio equal to or less than 2. In this instance, of course, the null hypothesis would be rejected.

A special case which possibly merits attention is that in which, at the specified confidence level, there is one and only one parametric proportion which will yield a critical ratio equal to the one postulated. Any other parametric proportion will yield a larger one and thus refute the null hypothesis. For example, if we set $p_2 = .57$, $p_1 = .43$, $N_1 = N_2 = 100$, $\rho = 0$, and $k = 2$, solution of the equation gives $p = .5$. Any value of p smaller or larger than .5 gives a k greater than 2.

The usefulness of the method proposed here is especially clear when the obtained proportions are near one end of the scale. Let $p_1 = .05$, $p_2 = .15$, $N_1 = N_2 = 100$, and $\rho = 0$. If we employ the conventional method, the estimated parametric proportion is .10, and, if this value is used, the critical ratio is found to be 2.358. However, this conclusion is bound to the value of the estimated parameter. There is nothing which can assure us that the estimate is, in fact, correct. Use of the quadratic method shows that if the parametric proportion were even as small as .146, our conclusions would require modification. Quite generally, in using the traditional method, the investigator proceeds on the assumption that the parametric proportion lies roughly mid-way between the two obtained proportions, and this usually is not likely to be the case.

---

[1] It is interesting to note in this connection that if the two roots of the equation are real numbers, their sum equals 1.

REFERENCE

1. LINDQUIST, E. F. A First Course in Statistics. New York: Houghton-Mifflin, 1942.

Department of Psychology
University of California
Los Angeles 24, California

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 131-135.

AN EXPERIMENTAL TEST OF AN HYPOTHESIS FOR PREDICTING OUTCOME WITH ELECTROSHOCK THERAPY*

Division of Psychiatry and School of Medicine, University of California

LEWIS G. CARPENTER, JR.[1]

A. THE PROBLEM

There are numerous theories (2) to account for the therapeutic effect of electroshock but the literature affords few examples of an experimental integration of theory and empirical fact. This paper presents an experimental approach to the major hypothesis that unfavorable outcome with electroshock therapy is related to externally directed aggression which is unacceptable to the ego. Fortunately the means for doing this were at hand in the publication of an empirically derived and validated scale of the Minnesota Multiphasic Personality Inventory (*MMPI*) which predicts outcome with shock therapy (1). It remained to determine the content of this scale with respect to the dimension of ego-alien externally directed aggression.

A hypothesis frequently made to account for improvement with shock therapy is that the provision of shock fulfills, at least temporarily, a present superego need (3). This hypothesis seems to have arisen jointly from the psychoanalytic theories of life and death wishes and the economy of the superego, and from empirical observations of a relatively high rate of remission in affective states following the application of shock.

A critical examination of the theoretical positions referred to above is not appropriate here except to remark that both require the active acceptance of self-criticism by the ego. According to the hypothesis inherent in these theories, it would be expected that ego syntonic internally directed aggression is positively associated with favorable outcome with shock, while its structural opposite, ego-alien externally directed aggression would be expected to be negatively associated with favorable outcome with shock.

Feldman (1) compared the response to *MMPI* items of successive groups of patients being considered for shock therapy in terms of favorable or unfavorable outcome, which was established later. In this way 52 items

---

*Received in the Editorial Office on February 5, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to express his appreciation to Dr. R. E. Harris and Dr. G. H. Crook for their valuable assistance.

were found significantly and dependably to discriminate the probability of unfavorable outcome. These items together form a shock prediction scale which separated 42 recovered patients from 42 unimproved cases at a probability level of 0.001 (critical ratio = 9:65).

The scale is derived so that it identifies those persons who have an unfavorable outcome with shock therapy and only indirectly those having the likelihood of a favorable outcome. It should be noted that the items of the scale were selected by an empirical process which disregarded the content of the items singly or as a scale. Except by rational analysis following this selection, a very difficult and uncertain matter, no statement of content or theoretical implications was attempted.

Two useful consequences would result from determining the content of the Shock Prediction Scale. First, the Scale could then be used as a measure for the determined variable. Second, exidence would be at hand relevant to the content of psychological variance associated with unfavorable outcome with shock. As a corollary there would be inferential evidence about the content of psychological variance associated with favorable outcome with shock.

B. THE EXPERIMENT

Following the theoretical line of thought above, it seemed plausible that the content of the Shock Prediction Scale, since high scores on it are associated with unfavorable outcome, might be ego-alien extrapunitive in content. In order to determine whether the content of the Shock Prediction Scale is in fact ego-alien extrapunitive, a sorting experiment was devised. The essentials of the experiment are as follows:

1. A second group of 52 items was selected by taking every ninth card from the total shuffled pack less the 52 items of the Shock Prediction Scale.
2. The combined shuffled pack of 104 items was presented to 20 psychologically sophisticated judges.
3. The judges were given the following instructions:

> *a.* Sort the cards into two categories: (*a*) relevant to ego-alien extrapunitiveness, (*b*) not relevant to ego-alien extrapunitiveness. File the relevant items in a single group in front of the box. File the irrelevant items at the rear of the box.
>
> *b.* Working with the ego-alien extrapunitive items only, file behind the appropriate separators: (*a*) items which ego-alien extrapunitive persons answer *true,* (*b*) items which ego-alien extrapunitive persons answer *false.* Make each judgment independently of all others. There may be any number of cards from all to none in any category. Make the selection in accordance with how the ego-alien extrapunitive per-

son would respond to the items and not indirectly according to what you judge the personality structure to be.

c. *Specification of ego-alien extrapunitiveness.* The items are to be judged ego-alien extrapunitive when aggression can be inferred and when alienation of hostility from the ego is employed as the defensive dynamism. This is distinguished from frank aggression. Hostility is not felt by the ego-alien extrapunitive person to belong to him. It is not integrated with the ego, and the person feels he has incomplete responsibility for the aggression. Following are some of the ways that ego-alien aggression is manifested: (*a*) a feeling of victimization, (*b*) unrealistic degree of kindliness, (*c*) displaced hostility, e.g., social movements, (*d*) sudden aggressive impulses whose origin is unknown to the subject; repression, (*e*) ingratiation to force supplies from the world; seduction, (*f*) sublimation, rationalization, denial, undoing, isolation of hostility; counter-cathectic dynamisms, (*g*) guilt following aggression, but provocation is cited.

4. After each judge had sorted the cards, two tabulations were made: (*a*) Those items which had been chosen to be relevant to ego-alien extrapunitiveness. (*b*) The direction (true or false) which the judge thought an ego-alien extrapunitive person would assign the item.

5. Statistical analysis was made of the items of the Shock Prediction Scale together with a comparison between it and the dummy scale; only relevance and direction when in agreement with the Shock Prediction Scale was considered.

C. RESULTS

The data from sorting can be analyzed in different ways to determine the connection between ego-alien extrapunitive personality structure and favorable or unfavorable outcome with shock therapy.

Let us first make a simple comparison between the number of Scale items and the number of dummy items which were judged to be relevant to ego-alien extrapunitiveness. Thirty-six (69 per cent) of the 52 Scale items were judged relevant by a simple majority of judges (11 or more out of 20), while 22 (42 per cent) of the 52 dummy items were so judged. This significant degree of separation is maintained when the criterion is raised to 15 or more of the 20 judges selecting an item to be relevant. Using this criterion, 25 (48 per cent) of the 52 Scale items were judged relevant, while only 10 (19 per cent) of the dummy items were so judged. Items which are predictive of unfavorable outcome with shock therapy are judged ego-alien extrapunitive in content.

Somewhat more sophisticated treatment is desirable. The mean number of judgments for relevance per item accorded the 52 Scale items was 13.0 with a standard deviation of 5.7. The mean number of judgments for relevance per item accorded the 52 dummy items were 9.6 with a standard deviation of 5.5. Comparing the difference between these distributions, the critical ratio is 3.14, yielding a probability of 0.002. That is, the chances are less than one in 500 that the difference is a chance difference.

The determination is superior to conventional confidence limits, and we may conclude that the shock prediction scale as a whole is ego-alien extrapunitive in content.

We may now raise a question refining the observed significant separation described above. The matter of relevance of item content has been established beyond reasonable doubt. The question is, are relevant items which are judged true for people with ego-alien extrapunitive personality structures also answered true by those who have an unfavorable outcome with shock therapy and vice versa. In other words, is the direction of response for ego-alien personalities consistent with the direction of response of patients having unfavorable outcome with shock therapy? This question can be approached by comparing, for the items relevant to ego-alien extrapunitiveness, the number of judgments made in the same direction with the number made in the opposite direction from the empirical key which predicts unfavorable outcome.

Following this procedure with the Scale items, the mean number of judgments in the same direction as is made by patients having unfavorable outcome is 8.6, with a standard deviation of 5.39. The mean number of judgments in the opposite direction as is made by patients having unfavorable outcome is 4.4, with a standard deviation of 3.29. The critical ratio for the difference between these means is 4.8, yielding a probability of 0.0001. That is, the chances are less than one in 10,000 that the obtained difference is a chance difference.

We may conclude from this that the indicated direction of response for persons with ego-alien extrapunitive personality structure is consistent with the observed direction taken by patients who have an unfavorable outcome with shock therapy.

D. Conclusions

The consensus of expert judges is that the shock prediction scale as a whole is ego-alien extrapunitive in content. This applies both to the scale as a whole and to agreement as to direction of scoring (true or false) between the judges and the empirically derived key of the scale. It follows that there is concordance between an ego-alien extrapunitive personality structure and poor prognosis with shock therapy. By inference, favorable outcome with shock therapy is probably related to the opposite personality structure, namely ego-syntonic internally directed aggression.

E. Summary

Twenty judges sorted 52 *MMPI* items (Feldman scale) predictive of unfavorable outcome with shock therapy, mixed with an equal number of randomly chosen items. Sorting was done for direction and relevance to a specification of externally directed aggression which is unacceptable to the ego. The shock prediction scale items were chosen significantly more frequently than the dummy items. There was better than chance agreement in direction.

One content variable of the shock prediction scale is ego-alien externally directed aggression. Ego-alien externally directed aggression is probably associated with unfavorable outcome with shock therapy. By theoretical analogy, ego-syntonic internally directed aggression may be associated with favorable outcome with shock therapy.

References

1. Feldman, M. J. A prognostic scale for shock therapy. *Psychol. Monog.*, 1951, **65**, No. 10.
2. Gordon, H. L. Fifty shock therapy theories. *Military Sur.*, 1948, **103**, 397-401.
3. Kalinowski, L. B., & Hoch, P. H. Shock Treatments and Other Somatic Procedures in Psychiatry. New York: Grune & Stratton, 1946.

Langley Porter Clinic
University of California Medical Center
Parnassus and First Avenues
San Francisco 22, California

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 137-141.

A STUDY OF MEDICAL STUDENTS WITH THE *MMPI*: II. GROUP AND INDIVIDUAL CHANGES AFTER TWO YEARS*

Department of Psychiatry and Neurology, University of Minnesota Medical School

WILLIAM SCHOFIELD

A. INTRODUCTION

During their junior year, University of Minnesota medical students spend 10 weeks on a clerkship on the psychiatry and neurology services. During this clerkship and in conjunction with a series of lectures on clinical psychology, each clerk takes the *MMPI*. In this way, a junior year profile was obtained for each of the male students whose freshman records were analyzed in the first paper of this series (5).

B. FRESHMAN-TO-JUNIOR *MMPI* CHANGES

Table 1 presents the mean scores on the *MMPI* clinical scales for three samples of male junior medical students and the corresponding codes for the mean profiles. The three samples are alike in showing highest mean elevation on the *Mf* scale and the second highest mean score on *Hy*. For all three, the mean profile is well within normal limits as measured from a *T*-score of 50. With respect to the neurotic triad, all three show the relationship which characterized the freshman mean profiles reported earlier, i.e., Hy>D>Hs. The average difference among the mean *T*-scores of the three samples, scale by scale, is only 2.6 points. When only the two larger samples are considered, the average discrepancy in mean *T*-scores is only 1.8 points, which is very similar to the average discrepancy obtained between the mean scale scores of the two freshman samples.

Table 2 reports the test-retest correlation coefficients derived from the freshman and junior *MMPI* records of the 83 male University of Minnesota medical students. On the clinical scales, these coefficients range from +.30 to +.59, indicating neither great stability nor great variability in the underlying measures. *These are not presented as reliability coefficients for the various scales* (2, 3, 4). The two-year interval is considerably longer than the period which intervenes in most instances where test-retest reliability is computed. When such an extensive period for the potential operation of

---

*Received in the Editorial Office on February 12, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

TABLE 1

MEAN SCORES ON THE *MMPI* CLINICAL SCALES FOR THREE SAMPLES OF MALE, JUNIOR MEDICAL STUDENTS AND CODES (1a) FOR THE MEAN PROFILES

Sample	*N*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*
Wisconsin (Blum (1))	25	50.0	50.8	54.9	51.4	58.7	49.9	50.3	52.8	49.5
Minnesota (Torrens (6))	52	43.7	48.9	53.5	48.2	55.0	51.3	46.2	45.8	49.7
Minnesota (Schofield)	83	42.8	49.3	54.5	49.8	57.4	49.6	42.0	43.6	51.6
		51.4*			57.8*			53.2*	55.8*	55.4*

Codes: Blum '53-
Torrens '5-
Schofield '53-18
non-K-corrected.

*K-corrected.

intervening experiences is added to the obviously basic variability of some of the personality components which the *MMPI* seeks to measure, it should be expected that test-retest coefficients will indicate greater fluctuation of scores than is acceptable when the usual "reliability" coefficient is being reported. It would appear that the medical students as a group are relatively more stable with respect to their activities, interests, and ideals (*Mf, Hy*) than in their somatic concerns and interpersonal conflicts or sensitivities (*Hs, Pd*). It is pertinent to note that the *Hy* scale, which yields one of the lower test-retest coefficients when reliability has been measured in general patient and normal groups (4), produces one of the highest retest coefficients in this sample which is characterized by a tendency to deviation on this scale.

TABLE 2

TEST-RETEST CORRELATION COEFFICIENTS FOR THE FRESHMAN AND JUNIOR YEAR *MMPI* SCORES OF 83 MALE MEDICAL STUDENTS

	Scale											
	L	*F*	*K*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*
"*r*"	.16	.46	.43	.41	.51	.52	.30	.59	.51	.41	.38	.32

In Table 3 are reported the freshman and junior mean scores for the sample of 83 male medical students. As indicated, statistically reliable changes in the mean scores were found on the three validity scales (*L, F,* and *K*), and two of the clinical scales (*Pd* and *Mf*). These changes suggest that over the two-year interval the students tend to become somewhat more sophisticated with regard to their ability to admit minor and relatively "acceptable" shortcomings (lower *L*) at the same time that they develop an increasing reserve (higher *K*) concerning more intimate aspects of their personal lives. Further, the data of Table 3 indicate a tendency toward relatively less femininity of interest and activity patterns (lower *Mf*) and greater capacity for acknowledgment of and open expression of hostilities toward "manifest authority" (higher *Pd*) after two years in the medical school. It is obvious that these shifts while reliable are very slight.

TABLE 3

FRESHMAN AND JUNIOR YEAR MEANS AND SIGMAS OF *MMPI* SCORES OF 83 MALE MEDICAL STUDENTS[a]

Scale	Freshman		Junior		*CR*	*CR*
	Mean	*SD*	Mean	*SD*	Means	Sigmas
L[b]	2.4	1.6	1.6	1.6	3.570**	0.000
F[b]	3.0	2.1	2.6	1.8	2.037*	1.685
K	59.8	7.7	63.9	6.5	5.026**	1.750
Hs	50.9	7.3	51.4	7.4	.556	.013
D	49.4	8.0	49.3	8.3	.162	.371
Hy	54.4	6.7	54.5	7.1	.224	1.157
Pd	54.1	7.9	57.8	8.7	3.432**	.958
Mf	59.4	9.7	57.4	8.4	2.078*	1.600
Pa	50.0	6.4	49.6	7.0	.456	.870
Pt	52.8	8.1	53.2	8.9	.318	.913
Sc	53.5	6.9	55.8	7.8	1.910	1.167
Ma	58.7	8.5	55.4	7.6	3.200**	.735

[a] Subjects were all the male members of a class of medical students tested, after admission to medical school, in the freshman year and again in the junior year.

[b] Raw score data.

*Significant at 5 per cent level.

**Significant at 1 per cent level.

A further analysis of the nature of the changes in the *MMPI* profiles of this sample of medical students over the two-year interval is shown in Table 4 which reports the frequency with which each of the scales was the high point in the freshman and junior profiles respectively. Extensive change in the frequency with which they appear as high points is shown by three scales: *Pd, Mf,* and *Ma*. From the freshman to the junior years the *Pd* scale shows an increase of over 14 per cent in the frequency with which it is the high point in profiles of this sample of medical students. The *Mf* and *Ma* scales show sizeable decreases in their frequency as high points. These changes are a further reflection of the personality trends suggested by the changes in the mean scores of the group.

TABLE 4
COMPARISON OF FREQUENCIES OF *MMPI* HIGH-POINT CODES IN FRESHMAN AND JUNIOR YEARS IN PER CENTS

Code	Scale	Freshman	Junior
Indeterminate*		10.9	6.0
1	*Hs*	4.8	1.8
2	*D*	2.4	4.2
3	*Hy*	7.2	6.0
4	*Pd*	5.4	19.8
5	*Mf*	32.5	27.4
6	*Pa*	1.8	2.4
7	*Pt*	6.0	4.2
8	*Sc*	4.2	7.8
9	*Ma*	22.3	16.8
No high-point**		2.5	3.6
		100.0	100.0

*A profile is coded as "indeterminate" if the three highest scale values are equal or within one point of each other.

**A profile with "no high point" is one in which no one of the clinical scores exceeds T-54.

C. SUMMARY AND CONCLUSONS

This paper, the second in a series of three, has reported the changes in personality associated with two years in a medical school curriculum, with data from freshman and junior year *MMPI* records furnishing the material for analysis. Freshman and junior profiles were available for the 83 male members of the University of Minnesota Medical School class of 1950.

A study of test-retest correlation coefficients, changes in mean profiles, and changes in frequency with which each *MMPI* scale occurred as the high point in individual profiles suggests the following generalizations:

1. It appears that the freshman medical students are not rigidly fixed as regards any of the personality components measured by the *MMPI*. They are apparently sensitive to certain kinds of experience and to maturing stresses, but the experiences and stresses of the first two years of medical school do not have equal and consistent effect on all aspects of their personalities. They tend to show individual stability with respect to their cheerful acceptance of deprivation and frustration, their expression of an optimistic evaluation of the future, and their professed belief in the goodness of people (*Hy*). Furthermore, as a group, they are not significantly different in these respects after two years.

2. Likewise, the medical students show high consistency relative to each other in the patterning of their interests, likes and dislikes avocationally and recreationally, and general deviation from their own sex norm in these

respects (*Mf*). However, while relatively stable in these respects as individuals, as a group they show a significant tendency to regress toward the mean, i.e., become more masculine psychologically, after two years. The shift, while slight, is sufficient to indicate a reliable change. The original distribution of scores was not sufficiently deviant to argue that the retest scores are accounted for solely by statistical regression to the mean.

3. From their freshman to junior years, the medical students show a significant increase in the frequency and extent to which they are able to express negative attitudes regarding their interpersonal relationships (*Pd*).

4. The enthusiasms, confidence, and general élan manifested in the freshman year show a very clear reduction by the junior year although, as a group, the medical students remain still above average in these respects (*Ma*).

The final paper in this series will report on the relationship between *MMPI* profiles and academic achievement in medical school.

REFERENCES

1. BLUM, I. P. A comparative study of students preparing for five selected professions including teaching. *J. Exper. Educ.*, 1947, **16**, 31-65.

1a. HATHAWAY, S. R. A coding system for *MMPI* profile classification. *J. Consult. Psychol.*, 1947, **11**, 334-337.

2. HATHAWAY, S. R., & MCKINLEY, J. C. A multiphasic personality schedule (Minnesota): III. The measurement of symptomatic depression. *J. of Psychol.*, 1942, **14**, 73-84.

3. MCKINLEY, J. C., & HATHAWAY, S. R. A multiphasic personality schedule (Minnesota): IV. Psychasthenia. *J. Appl. Psychol.*, 1942, **26**, 614-624.

4. ————. The Minnesota Multiphasic Personality Inventory: V. Hysteria, Hypomania, and Psychopathic Deviate. *J. Appl. Psychol.*, 1944, **28**, 153-174.

5. SCHOFIELD, W. A study of medical students with the *MMPI*: I. Scale norms and profile patterns. *J. of Psychol.*, 1953, **36**, 137-141.

6. TORRENS, J. K. An investigation and evaluation of the Guilford Inventory of Factors *STDCR* with special reference to the *MMPI*. (Unpublished paper, Univ. Minnesota, 1944.)

Department of Psychiatry and Neurology
University of Minnesota Medical School
Minneapolis 14, Minnesota

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 143-151.

STUDIES OF VESTIBULAR FUNCTION: I. THE DURATION OF PRIMARY NYSTAGMUS AS A FUNCTION OF SPEED OF ROTATION AND OF ACCELERATION*[1]

Department of Psychology, University of Rochester

G. T. Hauty[2] and G. R. Wendt[3]

A. Introduction

This is one of a series of reports of research on various aspects of vestibular function done at the University of Rochester. It describes a preliminary exploration of the relation of the duration of primary ocular nystagmus in man to the total amount of work done on the ear, i.e., speed of rotation, and to the rate of work done on the ear, i.e., rate of acceleration of rotation.

B. Procedures

The subject was seated on a rotating platform[4] with head fixed at a forward tilt of about 15°. The vertical axis of the head was at the rotation axis. The platform was driven by a power unit (located in the adjoining room for purposes of sound isolation) which consisted of a hydraulic transmission driven by a constant speed ¾ HP motor. This could turn the chair in either direction at any speed up to 180°/sec. Acceleration and deceleration were controlled either by starting and stopping the motor while the transmission was set for the desired speed, or by the rate at which the control lever of the transmission was driven with the motor running. In this study the accelerations used were approximately constant.

Each *stimulation period* consisted of an acceleration period followed by five minutes of rotation at constant speed and a deceleration period of the same character followed by five minutes of sitting still. There were thus two equal but opposite stimuli in each *stimulation period,* each followed by five minutes of observation. Three such periods, separated by rest periods

---

*Received in the Editorial Office on February 13, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] This work was supported in part by funds of the Office of Naval Research under a contract with the University of Rochester (Contract N6ori-126, Task Order 1).

[2] USAF School of Aviation Medicine, Randolph Field, Texas.

[3] Department of Psychology, University of Rochester.

[4] Originally built by Raymond Dodge and modified by G. R. Wendt with the assistance of G. T. Hauty and Gordon Gay.

of about 15 minutes, made up an experimental session, giving six stimuli per session. All rotation was to the subject's right.

For each subject the experiment can be divided into three parts: (*a*) A period of training in observation, made up of 15 or more *stimulation periods* using various speeds and accelerations. (*b*) Fifteen experimental sessions of three *stimulation periods* each during which speed of rotation was the experimental variable. The speeds used were 11, 22, 45, 90, and 180°/sec., given in random order. The accelerations and decelerations were brief and roughly equal in duration, achieved by closing and opening the circuit of the electric motor power source. (*c*) Four experimental sessions of three *stimulation periods* each during which rate of acceleration was the experimental variable, speed being always 180°/sec. The rates of acceleration used were such that the speed of 180°/sec. was reached in 0.5, 5, 30, 85, or 165 seconds. The same rate was used for acceleration and deceleration within each *stimulation period*, but was randomly varied between periods. The 0.5 sec. acceleration was done by starting and stopping the motor while the transmission was set for 180°/sec.; its form is therefore non-linear. The other accelerations were done by driving the control of the hydraulic transmission at a constant speed while the motor was running, thereby getting a nearly constant acceleration or deceleration.

The records consisted of verbal reports of the apparent movement of a fixation object, i.e., oculogyral movement. The room was completely dark. The subject looked at the fixation object, a collimated star loaned to us by Capt. Graybiel, identical with that used in the Tulane-Pensacola studies (5). The star was attached to the platform directly before the subject in the normal line of regard. This object is a pin-point of light with six radiating lines having the characteristic of being at optically infinite distance.

A description of the characteristics of oculogyral movement and of its relation to ocular nystagmus has been given elsewhere (9). The sequence of events under these conditions is an apparent movement of the star which is the result of the ocular nystagmus. As the subject begins to be rotated toward his right, the star appears to move to the right, since the apparent movement is the result of the movement of the images across the retinae during the slow phase of nystagmus which in this case is to the left. As nystagmus slows, the apparent movement slows, and then reverses its direction as the primary nystagmus is replaced by the secondary inverse nystagmus. The inverse gradually speeds up, then slows, and finally stops or is replaced by further reversals. In some cases there is a pause in the seen movement at the time when primary nystagmus slows and reverses into the secondary nystagmus. For the purposes of this paper only the duration of the movement during primary nystagmus was taken from the records.

The verbal report was recorded by a Sound Scriber, placed on the platform with its microphone just before the subject. The subject gave continuous reports of the direction of movement of the star and of its apparent speed. The training period emphasized the importance of exact report of the reversal of direction of movement when the primary nystagmus was replaced by the secondary inverse nystagmus, or, if apparent movement ceased between primary and secondary nystagmus, then exact reports of its cessation near the end of the primary and onset after beginning of the secondary nystagmus.

The durations of movement were measured by timing, from the Sound Scriber records, the elapsed time between onset of acceleration or deceleration and incidence of the cessation or the turning point in the direction of oculogyral movement.

The validity of the subject's report of cessation or turning of oculogyral movement, as a reasonably exact indicator of the duration of the primary nystagmus, is probably good enough for the purposes of the present study. Other studies in this laboratory have shown that such durations correspond closely to eye-movement records when eye movements and subject's report are recorded concurrently. However, their validity depends on the training and skill of the subject. It is likely that the relatively untrained subject often reports cessation of movement, or withholds a report of cessation because he feels uncertain, or that autokinetic movement, not based on nystagmus, is experienced. At the end of this series of observations, when each subject had given 74 or more reports, the reported times showed a remarkable consistency, with only an occasional out-of-line report. Each subject was serving also in another study of oculogyral movement under linear accelerations, thus getting added practice. Nevertheless, some invalid and other slightly inaccurate reports are undoubtedly included in the data. (These may be identified by inspection of deviant times in Table 1.) Their overall effect is probably to increase slightly the times in the first part of the study, that of the effects of speed rotation. This follows from the fact that out-of-line and invalid reports are limited at the low end (shortest reported duration = 14 sec.), but not at the high end (longest reported duration after brief acceleration = 69 secs.).

There were three subjects. Their identity is named so that one can make comparisons with possible later reports on the same persons. Subject *A* was Patricia Dougan, age 23; *B* was George Hauty, age 30; *C* was Carson Johnson, age 40. The study was done in July of 1949.

C. Results

Table 1 presents the raw data for each of the three subjects for each of the 54 stimuli. These are presented as an aid to interpretation of results by making possible identification of deviant times. Each individual cell shows the duration in seconds of the oculogyral movement of primary nystagmus from a single stimulus, with columns arranged to separate the three subjects and to separate the accelerations and decelerations. The rows are arranged to separate the nature of the stimulus and the three *stimulation periods* for each stimulus. The upper part of the table gives that part of

TABLE 1
Duration of Oculogyral Movement of Primary Nystagmus

Stimulus conditions	Stim. Period	Raw data of 3 *S*'s					
		Subject *A*		Subject *B*		Subject *C*	
		Accel.	Decel.	Accel.	Decel.	Accel.	Decel.
Veloc. 11°/sec.	1	36 secs.	28 secs.	30 secs.	18 secs.	69 secs.	17 secs.
Accel. 0.5 sec.	2	33	35	43	20	21	16
	3	41	22	15	32	19	17
Veloc. 22°/sec.	1	29	16	30	26	41	15
Accel. 0.5 sec.	2	29	34	21	20	21	17
	3	35	35	15	21	19	17
Veloc. 45°/sec.	1	42	26	50	25	22	23
Accel. 0.5 sec.	2	25	28	29	26	22	22
	3	24	25	17	28	21	22
Veloc. 90°/sec.	1	29	42	33	27	23	25
Accel. 0.5 sec.	2	34	39	27	26	26	21
	3	24	31	25	27	19	17
Veloc. 180°/sec.	1	32	34	30	26	22	21
Accel. 0.5 sec.	2	33	29	33	29	20	23
	3	31	21	33	27	14	19
Veloc. 180°/sec.	1	35	33	34	28	21	26
Accel. 5 sec.	2	37	31	29	36	17	18
	3	47	32	29	28	23	21
Veloc. 180°/sec.	1	64	60	51	48	40	42
Accel. 30 sec.	2	59	46	50	47	32	35
	3	71	53	48	44	42	45
Veloc. 180°/sec.	1	95	44	94	94	71	88
Accel. 85 sec.	2	109	96	95	90	79	86
	3	92	105	91	93	52	86
Veloc. 180°/sec.	1	162	175	164	164	150	164
Accel. 165 sec	2	204	174	165	170	125	150
	3	164	173	162	172	148	163

the experiment in which velocity was varied, the lower part that in which acceleration was varied.

TABLE 2

MEAN DURATION OF NYSTAGMUS DURING ACCELERATION, DECELERATION AND ACCORDING TO STIMULATION PERIOD

	Time in seconds			
Conditions	Subject *A*	Subject *B*	Subject *C*	Mean
All accelerations, Study I	32 secs.	29 secs.	25 secs.	28 secs.
All decelerations, Study I	30	25	20	25
All accelerations, Study II	95	84	67	82
All decelerations, Study II	85	85	77	82
All *stimulus periods* 1, Study I	31	30	28	30
All *stimulus periods* 2, Study I	32	27	21	27
All *stimulus periods* 3, Study I	29	24	18	24
All *stimulus periods* 1, Study II	84	85	75	81
All *stimulus periods* 2, Study II	95	85	68	83
All *stimulus periods* 3, Study II	92	84	73	83

Table 2 shows a summary of the data by means, broken down to show separately the duration of nystagmus for acceleration and for deceleration and for the three *stimulation periods* used with each stimulus value. These are presented to enable one to assess the rôle of a number of possible factors such as vestibular imbalance, training, or habituation. "Study I" and "Study II" refer to the data in the upper and lower parts respectively of Table 1. "All *stimulus periods* 1" includes durations of both acceleration and deceleration nystagmus for the first presentation of all stimulus values; "2" is the second presentation; "3" the third. Means of the three subjects are shown in the last column. Table 2 shows that in Study I there was a decrease in mean duration of oculogyral movement during the experiment from 30 secs. for all *stimulation periods* 1, to 24 seconds for all *stimulation periods* 3, and a decrease from the accelerations to the decelerations from 28 to 25 seconds. In Study II, however, neither of these differences appears in the means of the group. The decreases found in Study I are therefore attributed to the effects of practice in reducing spuriously long reports of apparent movement, rather than to habituation to rotation (9). The second and third *stimulation periods* occurred in the second and last thirds of the time taken by the study; the decelerations occurred five minutes after the subject had reported on a similar acceleration; hence in each case there was the advantage of practice.

TABLE 3
MEAN DURATION OF PRIMARY NYSTAGMUS AS A FUNCTION OF SPEED OF ROTATION

	Time in seconds Subject *A*	Subject *B*	Subject *C*	Mean
Vel. 11°/sec., accel. 0.5 secs.	32 secs.	26 secs.	27 secs.	28 secs.
Vel. 22°/sec., accel. 0.5 secs.	29	22	22	25
Vel. 45°/sec., accel. 0.5 secs.	28	29	22	27
Vel. 90°/sec., accel. 0.5 secs.	33	28	22	28
Vel. 180°/sec., accel. 0.5 secs.	30	29	20	27

Table 3 displays the relation of duration of nystagmus to speed of rotation. The results are substantially the same for each subject; there is no evident trend in the durations in going from weak stimuli to strong stimuli. According to the overall means and to the data of each subject, duration of primary nystagmus is not affected by strength of stimulus. Increase in stimulus strength affects only the rate of the slow phase of nystagmus and not its duration.

TABLE 4
MEAN DURATION OF PRIMARY NYSTAGMUS AS A FUNCTION OF DURATION OF ACCELERATION

	Time in seconds Subject *A*	Subject *B*	Subject *C*	Mean
Accel. 0.5 secs., vel. 180°/sec.	30 secs.	29 secs.	20 secs.	26 secs.
Accel. 5 secs., vel. 180°/sec.	36	31	21	29
Accel. 30 secs., vel. 180°/sec.	59	48	39	49
Accel. 85 secs., vel. 180°/sec.	90	93	77	87
Accel. 165 secs., vel. 180°/sec.	175	166	150	164

Table 4 presents the relation of duration of nystagmus to duration of acceleration, when the terminal speed is in each case rotation at 180°/sec. Inspection of these means shows that under the conditions of this experiment duration of nystagmus increased with increase in duration of acceleration. With small durations of acceleration the increase in nystagmus time probably corresponds nearly to the increased duration of acceleration. With longer durations, the persistence of nystagmus after the end of acceleration becomes less. With very long durations nystagmus possibly reverses its direction while the platform is still being accelerated. Reference to Table 1 shows that this always happened with the 165-second acceleration for subject *C*.

The data in the first row of Table 4 were taken from Study I, those from the last four rows from Study II. It is possible, therefore, as noted above, that the overall mean duration in the first row is spuriously high. Perhaps one should subtract about three seconds from this figure.

An analysis of variance was done for each of the two studies, using the data from Table 1 (omitting one highly deviant measure, the 69-second figure for Subject *C* with a velocity of 11°/sec.). This justified the following statements:

1. There are no significant differences in duration of primary nystagmus which may be attributed to the different velocities of rotation used to Study I.
2. There are significant differences in duration of primary nystagmus which may be attributed to the different rates (durations) of acceleration used in Study II (*t*-tests show that each of the rates yields a different duration, significant at better than the 1 per cent level).
3. There are no significant differences in duration of primary nystagmus between that following acceleration and that following deceleration.
4. There is significant inter-individual variation.

D. Discussion

The first result of this study, that low intensities of stimulation as measured by speed of rotation yield nystagmus of the same duration as high intensities, is not new. Buys (1) showed this to be so in 1924 on the basis of graphic records of eye movements taken through closed lids, thereby excluding visual stimulation. More recent investigations (8 for example) have reported that weak stimuli yield a much briefer nystagmus than strong stimuli. However, we are inclined to attribute their result to the presence of visual stimuli strong enough to inhibit the weak nystagmus following weak stimuli (9, p. 1201). In the present experiment the star was present, but the results testify to the fact that so weak a visual stimulus does not inhibit nystagmus significantly. Our results confirm Buys.[5]

This result has a bearing on the problem of what happens at the cupula during rotation. Since Steinhausen (7) and Dohlman (3, 4) directly observed the cupula in fish and since Lowenstein and Sand (6) recorded action potentials from the vestibular nerve of the ray, the conclusion has gained wide acceptance that the duration of primary nystagmus is determined by the duration of deviation of the cupula. Our data seem to cast doubt on this conclusion as being applicable to man. If we grant that the elastic restoring force of the cupula is small and that there is a high degree of damping because of the small bore of the canals, then the small cupular deviations produced by weak stimuli should be restored much more rapidly than the large deviations produced by strong stimuli. Our eye movement data would require that the restoration time be the same irrespective of the extent of the deviation. This suggests that nystagmus is not a continuous measure of cupular deviation and that it must be otherwise determined.

---

[5] A forthcoming paper by G. T. Hauty, No. II in this series, will report repetition of the present experiment with photographic recording of eye movements.

The same conclusion is also supported by the data of the second study, i.e., those showing that after long periods of acceleration (85 seconds, 165 seconds) nystagmus ends shortly after acceleration ends and possibly before it ends. [Dodge (2) reported this same finding in 1923.] Yet the restoration of the cupula to its position of rest probably takes as long after cessation of a long acceleration as it does after a short acceleration. If nystagmus were a measure of cupular displacement, the nystagmus should have lasted on the average 111 seconds as a result of the 85-second acceleration (arrived at by adding the 26-second duration obtained with accelerations of 0.5 sec. to the 85-sec. duration of acceleration) rather than the obtained 87 seconds. It should have lasted 191 seconds with the 165-second acceleration rather than the obtained 164 seconds. The cessation of nystagmus came 2, 7, 24, and 27 seconds earlier than it should have, with accelerations lasting 5, 30, 85, and 165 seconds, if the cupular deviation theory were true.

E. SUMMARY

This is one of a series of reports of research on various aspects of vestibular function done at the University of Rochester. It describes a preliminary exploration of the relation of the duration of primary ocular nystagmus in man to the total amount of work done on the ear, i.e., speed of rotation, and to the rate of work done on the ear, i.e., rate of acceleration of rotation. In the first part of the experiment each of three subjects was stimulated six times by acceleration to or from each of five speeds of rotation (11, 22, 45, 90, and 180°/sec.), with duration of acceleration constant at about ½ sec. In the second part of the experiment the same subjects were stimulated six times by each of five rates of acceleration to or from a speed in each case of 180°/sec. The rates were such that the speed of 180°/sec. was reached in 0.5, 5, 30, 85, and 165 secs. Duration of the primary nystagmus was determined by the method of oculogyral movement of a fixation object. In the first part of the experiment it was found that the duration of nystagmus was probably constant, unaffected by the amount of work done on the ear. Speeds of 11, 22, 45, 90, and 180°/sec. produced primary

nystagmus with mean durations of 28, 25, 27, 28, and 27 secs. respectively. On the other hand, in the second part of the experiment when the period of acceleration was lengthened, the primary nystagmus was also lengthened. With a fixed speed of 180°/sec., reached or stopped in various times, the duration of primary nystagmus was: accel. 0.5 secs., 27 sec. of nystagmus; accel. 5 secs., 29 sec. of nystagmus; accel. 30 secs., 48 sec. of nystagmus; accel. 85 secs., 87 sec. of nystagmus; accel. 165 secs., 164 sec. of nystagmus. With the two longest accelerations the data indicated that the primary nystagmus sometimes ended and was replaced by the secondary inverse nystagmus while subject was still being accelerated. Arguments are presented to show that the results are incompatible with attributing the duration of nystagmus to the duration of deviation of the cupula.

REFERENCES

1. Buys, E. Contribution à l'étude du nystagmus oculaire de la rotation chez l'homme. *Rev. d'oto-neuro-ocul.*, 1924-25, **2**, 641-659, 721-749; **3**, 10-32, 105-126.
2. Dodge, R. Habituation to rotation. *J. Exp. Psychol.*, 1923, **6**, 1-35.
3. Dohlman, G. The rôle of the perilymph in vestibular reactions. *Arch. Ohr., Nas., u. Kehlk-Heilk.*, 1941, **150**, 25-30.
4. ———. Investigations in the function of the semicircular canals. *Acto oto-laryng., Stockh.*, 1944, Suppl. 51, 211-219.
5. Graybiel, A., Clark, B. MacCorquodale, K., & Hupp, D. I. Rôle of vestibular nysagmus in visual perception of moving target in dark. *Amer. J. Psychol.*, 1946, **59**, 259-266.
6. Löwenstein, O., & Sand, A. The individual and integrated activity of the semicircular canals of the elasmobranch labyrinth. *J. Physiol.*, 1940, **99**, 89-101.
7. Steinhausen, W. Ueber die Beobachtung der Cupula in den Bogengangsampullen des Labyrinthes des lebenden Hechts. *Pflüg. Arch. Physiol.*, 1933, **232**, 500-512.
8. Van Egmund, A. A. J., Groen, J. J., & Jongkees, L. B. W. The mechanics of the semicircular canal. *J. Physiol.*, 1949, **110**, 1-17.
9. Wendt, G. R. Vestibular functions. In S. S. Stevens (*Ed.*) *Handbook of Experimental Psychology*. New York: Wiley, 1951. (Pp. 1191-1223.)

Department of Psychology
University of Rochester
Rochester 3, New York

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 153-163.

INTERRUPTION: ITS EFFECT UPON PERFORMANCE IN A "TROUBLESHOOTING" SITUATION*

Institute of Educational Research, Indiana University

NICHOLAS A. FATTU AND E. VICTOR MECH[1]

A. THE PROBLEM

There is, no doubt, some agreement that the process of interrupting a task constitutes an introduction of new factors in the stimulus situation. It is surprising, therefore, that so little effort has been made to determine experimentally the effects of interruption upon performance in operationally defined situations. A basic problem has grown out of this lack of systematic information.

The periodic assessment of the proficiency level of personnel with regard to performance on a particular equipment-system poses a problem not only for the armed forces but for industry as well. For example, the assessment of "troubleshooting" proficiency entails obtaining an equipment-system that the person works with on-the-job, and, under standardized conditions inserting problems or malfunctions for the man to locate.

In many instances the simulated situation will, at best, only approximate the actual one that confronts the maintenance man or troubleshooter. The physical equipment, for example, may be identical in two instances. However, in one case its location might be some intricate part of an aircraft while on the other hand, for the purpose of testing a group of mechanics, its location might be a laboratory room or a mobile unit. Clearly, in the latter case the stimulus properties might not be highly related to the situation a mechanic is actually faced with on-the-job.

It is well known that limitations of such approximations constitute a basic problem for psychologists. If the rationale of a performance test is to determine the proficiency level of a troubleshooter under standardized conditions, how, then, is this best accomplished? Various procedures with regard to accomplishing this are open to observers or examiners, as follows:

Procedure I. The subject is interrupted in the preceding task as well as in the present task in order for the examiner to obtain predetermined information.

---

*Received in the Editorial Office on February 14, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The authors are indebted to Mr. Walter Ciszczon for material contributions to apparatus, and to Mr. James Miller for assisting in the collection of the data.

Procedure II. The subject is permitted to complete the preceding task as well as the present task, but is "interrupted" by the examiner after each task for the purpose of obtaining information.

Procedure III. The subject is permitted to complete the preceding task as well as the present task, the examiner never attempting to obtain any verbal information from the subject, simply recording whether the malfunction is located and in what time limit.

Clearly, Procedures I and II provide the basic information that Procedure III does, and allow, also, the examiner to obtain extra data by virtue of questioning the subject as to knowledge possessed about the equipment.

It appears plausible to postulate that since a proficiency test would require repeated trials during which a considerable amount of learning could take place, performance might, in part, well be a function of the procedure used by the examiner to obtain information.

At present, however, the evidence is anything but conclusive with regard to whether interruption, from an operational point of view, produces cues that tend to engender performance in a problem situation.

These possible cues, coupled with the additional ones that examiners might provide during the course of asking questions, could temporarily make a mechanic appear more "proficient" than actually was the case. Clearly in a functional maintenance situation the troubleshooter does not, generally, have the benefit of these examiner cues.

That experimenters in the area of learning often are likely to underestimate their rôles in the situation is known well. The responsibility involved is embodied in Melton's (8) statement:

> Stimulus response and field theories frequently engage in the same types of investigations, with similar situations, but with such conditions imposed as to maximize the applicability of their own conceptual structure or minimize the applicability by the opponent's conceptual structure.

Brogden (5) points out, also, that the experimenter must play a large and frequently uncontrolled rôle.

Boguslavsky (4) describes a situation analogous to Procedures I and II that were referred to previously. His data showed that although the differences were not sufficiently significant to warrant the rejection of the null hypothesis, the trend was in favor of completed tasks to be recalled more often than the interrupted tasks. Hays (7) however, presents evidence that completed tasks are recalled more frequently when the interpolated task is complex and that interrupted tasks are more frequently recalled when they precede simple tasks. He adds that if the two types of tasks are combined in one experiment, results in either direction, or results showing no advantage for either condition, might be found.

The problem, then, is whether requiring a troubleshooter to synthesize and verbalize information under Procedures I or II produces differential results when compared with another group not receiving an opportunity to verbalize.

B. METHOD

1. *Apparatus*

The apparatus in Figure 1 is called a gear-train, consisting simply of a set of gears and shafts mounted on a piece of aluminum ½ inch thick, 24 inches in length, and 20 inches in width. The gear-trains were arranged to form two series and four parallel channels, which provided for crossed information chains. Two operating controls, *A* and *B*, provided the input necessary to obtain the desired motion. The motion was transferred through the gear-trains and as an end result closed a switch that caused a series

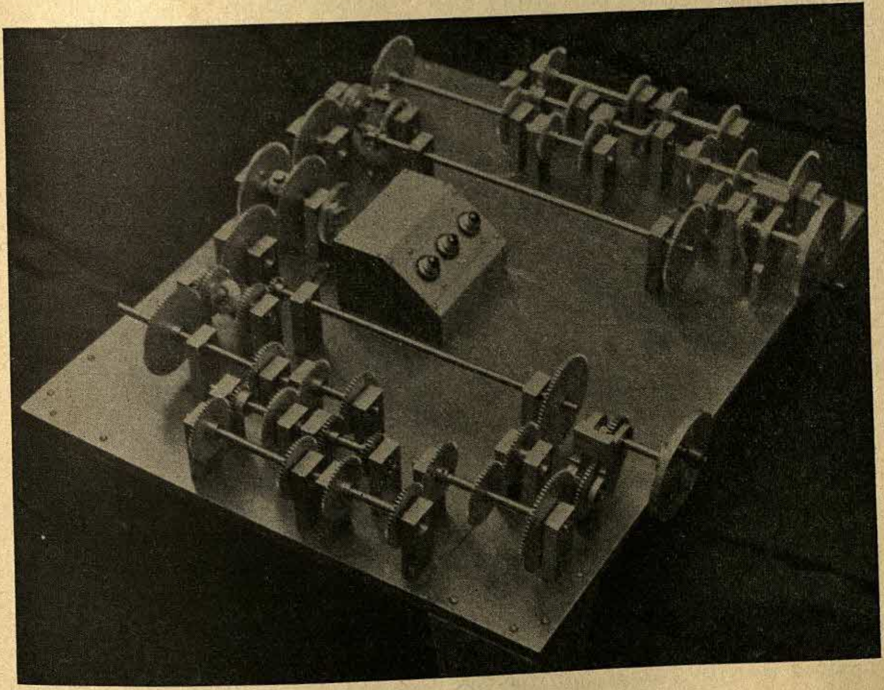


FIGURE 1
THE GEAR-TRAIN APPARATUS

of red lights to illuminate the control panel. The red lights would illuminate only if the equipment were working properly and Control *A* was turned 13 times and Control *B*, 12 times. When the appropriate number of turns was made and the expected end result (control panel lighting up) did not occur, this indicated to *S* that there was a "malfunction" in the gear-train.

2. *Malfunctions*

For the purposes of this experiment only three classes of "malfunctions" were utilized: (*a*) a gear not meshing, such as a distance between gears, (*b*) a slipping gear due to a set screw having been removed, (*c*) a damaged gear tooth. Each *S* received one malfunction from each class on the pre-test and one malfunction from each class on the post-test, making six malfunctions that each *S* was required to locate. *S*s were presented malfunctions in a random order, and, in addition, the location of each of six malfunctions on the gear-train was selected at random.

3. *Procedure*

The different conditions for the three groups are described in Table 1. However, certain training factors were common to all experimental groups.

*S*s were run individually, and immediately upon entering the laboratory for the first time, *E* gave *S* the Standard Operating Procedure (hereafter referred to as *SOP*) for the apparatus. The *SOP* simply consisted of turning the control handle *A*, 13 turns and control handle *B*, 12 turns, and if the red lights on the control panel did not light up, this indicated to *S* that something was wrong with the gear-train and the task was to "trouble-shoot" the equipment.

After the *SOP* orientation, *S* was given three malfunctions or problems to locate. Each problem was given singly, *S* being placed in an adjoining room while each new malfunction was being inserted by *E*.

After *S* had performed on the first three problems which were given to determine initial levels of ability, a tape recorded "basic knowledge" lecture lasting 10 minutes was given. Integrated with the lecture was a series of slides projected on a screen with a 35 mm. camera. The rationale of the lecture was to convey the basic nomenclature of the gear-train. Included were the concept of transfer of motion, function of gears, bearings, and shafts. *S*s received the lecture individually, which was played twice.

Up to this point, conditions were the same for all *S*s. They had all received the *SOP*, three initial problems, and a tape recorded basic knowledge lecture, played twice.

TABLE 1

SUMMARY OF EXPERIMENTAL PROCEDURES

Group	Pre-test	Training	Test conditions
I. Interrupted during troubleshooting.	All *S*s were given the Standard Operating Procedure and received three initial problems on the gear train.	All *S*s received a tape recorded basic knowledge lecture on transfer of motion, gears and shafts, and bearings.	Given three problems. Immediately after completing the *SOP* for each defect, *S* was required to indicate in which part of the equipment the defect was located.
II. Interrupted after defect was located.			Given three problems. After a decision was reached as to the location of the defect *S* was required to verbalize how the decision was reached in terms of information received from the equipment.
III. Control			Given three problems.

In the post-test, three malfunctions also were given to each *S* and the following procedure was used with the experimental groups:

Group I. As soon as the *S*s in this group had completed the *SOP* and the control panel did not illuminate, indicating to *S* that the apparatus was defective, *E* would interrupt with the following statement: "On the basis of the information you have received thus far from the apparatus, and without manipulating the controls for the moment, show me or tell me the area where you think the defect may be located."

Group II. When *S* had completed the *SOP* and searched for the malfunction and made a *decision* with regard to location of the malfunction, *E* would say: "Tell me how you decided that the defect is there. Try to tell me how you made this decision in terms of the checks you made and things noticed or information gotten from the equipment."

Group III was simply given three malfunctions to locate, and was not interrupted during the trouble-shooting process or after it.

*S*s were allowed no more than 15 minutes to locate each defect either on pre-test or test conditions. No knowledge of results was given to any *S*.

4. *Subjects*

The *S*s were 27 females enrolled in an educational psychology course in the School of Education at Indiana University. Three groups of nine *S*s each were used in the experiment. Assignment of the 27 *S*s to each of the three groups was done from a table of random numbers. Participation in the experiment was required in order to eliminate the bias often found by asking for volunteers.

C. RESULTS AND DISCUSSION

Figure 2 shows the number of *S*s in the three groups that correctly located each malfunction under initial and test conditions. Clearly, there were several inversions for the groups on the pre-test malfunctions, while under test conditions performance appeared more stable with respect to maintaining rank position.

To test whether statistically significant gains had been made by any group it was decided to take the difference between frequency of defects correctly located on the initial and test conditions. Theoretically, this procedure would provide seven number possibilities ranging from +3 to —3, permitting the use of a simple analysis of variance.

Prior to applying the analysis of variance, Bartlett's χ^2 test (2) was utilized to test the homogeneity of variance of defects located. Since the χ^2 value

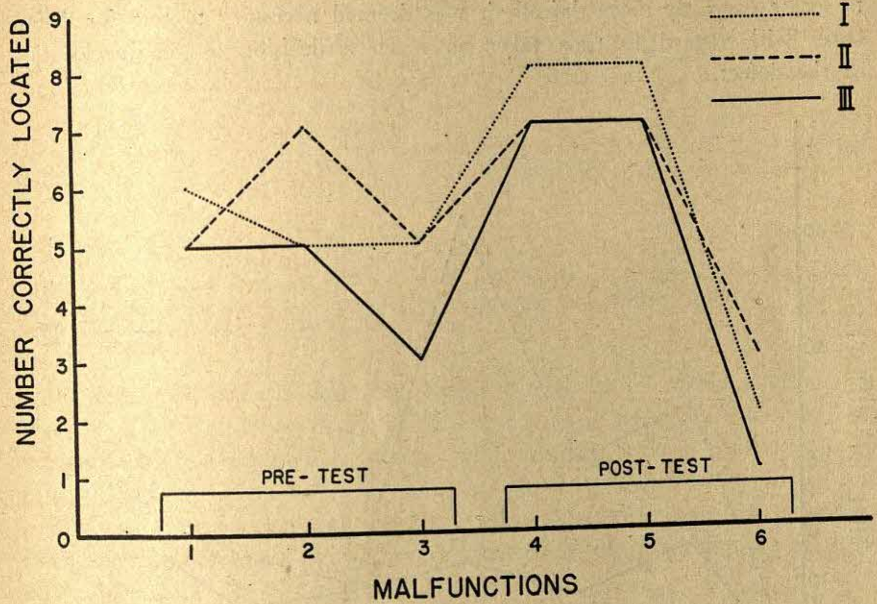


FIGURE 2
Frequency of Correctly Locating Each of Six Malfunctions for Three Groups of *S*s

of 2.15 did not reach the 5 per cent level ($df = 2$), the hypothesis of no differences among group variances could not be rejected.

The analysis of variance performed on the difference scores is presented in Table 2.

The obtained F value of 8.75 for 2 and 24 degrees of freedom falls far short of the value of 19.45 required for significance at the 5 per cent level.

TABLE 2
Analysis of Variance of Difference Scores

Source of variation	*df*	Mean square	*F*
Between groups	2	.148	8.75
Within groups	24	1.296	

Figure 3 shows the time in minutes required for the groups to reach a decision with respect to where the malfunction was located. Although the "interruptions" did not exert a differential effect, causing any one group to locate correctly more defects, it was deemed necessary to test for differences with regard to time taken to reach a decision as to the location of the defect.

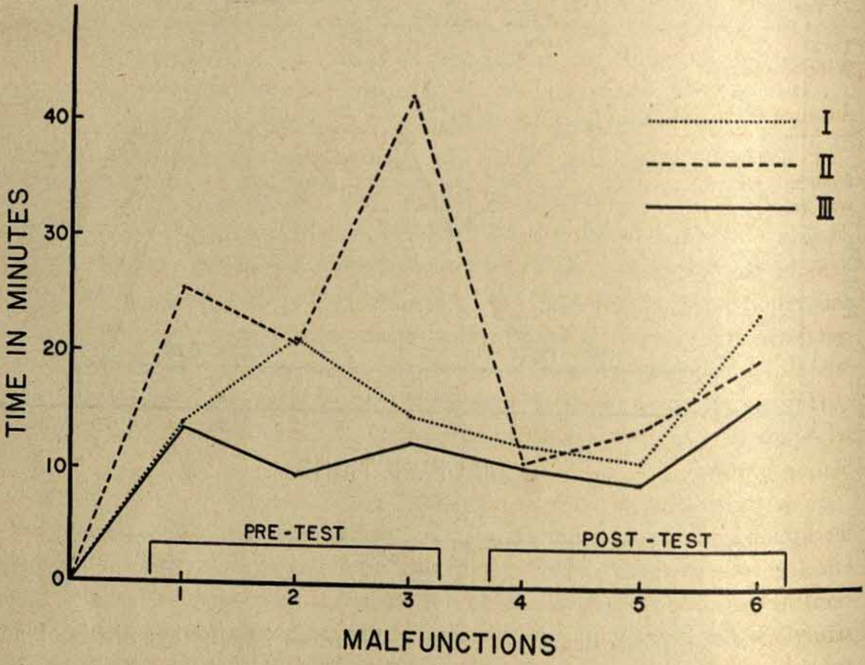


FIGURE 3
TIME REQUIRED FOR THREE GROUPS OF *S*s TO REACH A DECISION AS TO THE LOCATION OF EACH OF SIX MALFUNCTIONS

In order to reduce the heterogeneity of variance and correct somewhat for a skewed distribution, it was decided to apply a square root transformation to the time measures as recommended by Bartlett (1). An analysis of covariance, shown in Table 3, was carried out between the transformed pre-test and post-test total time measures.

The *F* of 0.00 is not significant, indicating that the means of the groups on the post-test time measures can be accounted for by differences in mean level of initial ability as measured in the pre-test trials. In effect, the experimental operation of interruptions, one group after the *SOP,* and another after they had completed one problem, and before they were given another, appears to produce no differential results in time taken to decide the location of a defect.

TABLE 3

ANALYSIS OF COVARIANCE OF THE $\sqrt{X}$ TRANSFORMED SCORES FOR TIME TO REACH A DECISION AS TO THE LOCATION OF SIX "MALFUNCTIONS"

Source of variation	Sum of squares of errors of estimate	*df*	Mean square	*F*
Total	13.1	23*		
Within groups	13.1	21*	.62	
Adjusted means	0.0	2	0.00	0.00

*Two degrees of freedom are lost since two missing time scores were estimated from the remaining data. A formula recommended by Cochran and Cox (6, p. 99) was employed to estimate the missing data.

Since the experiment described herein can be more correctly classed as one requiring problem-solving ability rather than one requiring recall of specific materials previously learned, direct comparisons with the data of others is not feasible.

However, certain general aspects of this experiment may be compared with other experiments dealing with the effects of interruption and completion on recall.

In a theoretical article Boguslavsky used a procedure in which the interruption of a task is simultaneous with the presentation of a new task, but the subject fails to respond to the new task immediately. This situation, he points out, is likely to occur when the experimenter gives oral instructions for a new task, at the same time that the original task was interrupted. Supposedly a change from visual to auditory stimulation would cause a change in postural set and receptor orientation.

In the present situation some *S*s were orally instructed to stop working on the gear-train after the *SOP* was completed and required to verbalize information gained up to that point, while another group was allowed to complete each problem before being required to verbalize the necessary information. Interestingly enough, also using oral instructions, Boguslavsky and Guthrie (3) found no statistically significant differences between two groups in recall, one of which was permitted to complete each task while the other was interrupted during each task.

The results obtained from the present experiment are similar to those reported by Boguslavsky and Guthrie. They found no significant differences in effects between subjects permitted to complete each problem and those who were interrupted while working on each problem. Similarly, for the operations employed in this experiment, no differences were apparent between the three groups for number of malfunctions solved or time to reach a decision.

The results, however, should be reviewed as provocative of further basic experimentation. Evaluating the trouble shooting proficiency level of a maintenance man often entails interruption and, consequently, the asking of questions while in the actual process of trouble shooting.

Although for the equipment, task, and procedure used in the experiment being reported, no differential effects were found, one should be cautious in assuming that these interruptions do not necessarily have effects.

There are several explanations with respect to the conditions necessary for interruption to have a differential effect. For example, the results obtained by Hays indicate that if simple and complex tasks are *combined* in one experiment, results in either direction might be found.

Admittedly, this was the case with the gear-train malfunctions. The task of locating the defect in which slippage was caused due to the removal of a set screw was clearly more complex than locating two gears separated, or a broken gear tooth.

In terms of the evidence presented here, it is to be concluded, then, that no advantage exists for "schedules" of interruption when problems or malfunctions of varying complexity are combined.

D. SUMMARY

Twenty-seven female *S*s participated in an experiment on the effects of interruption in a trouble-shooting situation.

Certain training factors were common to the three groups. All *S*s received thorough indoctrination in the Standard Operating Procedure for a gear-train apparatus, after which each *S* was given three problems or defects to locate in the equipment. This procedure was used to obtain a pre-test measure of "trouble shooting" ability on the gear-train. After the initial measures each *S* received a 10-minute basic knowledge lecture that was played twice. The lecture discussed the nomenclature and functioning of the gear-train. Finally, under test conditions each *S* was given three gear-train malfunctions to locate.

In addition, however, one group was interrupted immediately upon completion of the *SOP* and required to verbalize where the defect was thought to be located. This group was also asked what information was obtained from the *SOP* that helped *S* to arrive at a decision. After *S* had given the required information the instructions were to continue searching for the defect until a definite decision could be reached. There was, however, a 15-minute time limit on each defect for all *S*s.

Another group was allowed to complete each task, and then, *S*s were required to verbalize how they had located the defect, what information was obtained from the equipment, and what checks were made. The third group was allowed to complete each task without being questioned by the experimenter.

No differential effects in troubleshooting performance were found that could be attributed to interruption or completion. These data suggest that if simple and complex tasks are combined, results showing no advantage for either condition will be found.

REFERENCES

1. BARTLETT, M. S. Square root transformation in analysis of variance. *J. R. Statist. Soc. Suppl.,* 1936, **3**, 68-78.
2. ———. Some examples of statistical methods of research in agriculture and applied biology. *J. R. Statist. Soc. Suppl.,* 1937, **4**, 137-170.
3. BOGUSLAVSKY, G. W., & GUTHRIE, E. R. The recall of completed and interrupted activities: an investigation of Zeigarnik's experiment. *Psychol. Bull.,* 1941, **38**, 575-576 (abstract).
4. BOGUSLAVSKY, G. W. Interruption and learning. *Psychol. Rev.,* 1951, **58**, 248-255.
5. BROGDEN, W. J. Animal studies of learning. In S. S. Stevens (*Ed.*), *Handbook of Experimental Psychology.* New York: Wiley, 1951.
6. COCHRAN, W. G., & COX, G. M. Experimental Designs. New York: Wiley, 1950. Pp. IX+454.
7. HAYS, D. G. Situational instructions and task order in recall for completed and interrupted tasks. *J. Exper. Psychol.,* 1952, **44**, 434-437.
8. MELTON, A. W. Learning. In C. P. Stone (*Ed.*), *Annual Review of Psychology.* Stanford, Calif.: Annual Reviews, Inc., 1950. (Pp. 9-30.)

Institute of Educational Research
Indiana University
Bloomington, Indiana

Published as a separate and in *The Journal of Psychology*, 1953, 36, 165-186.

PARTIAL VERBAL REINFORCEMENT RELATED TO DISTRIBUTED PRACTICE IN A CLASSROOM SITUATION*

Institute of Educational Research, Indiana University

Donavon Auble and E. Victor Mech[1]

A. Introduction

That educational psychologists should, in the main, concern themselves with experimentation in the area of human learning, is repeatedly pointed out in current textbooks. Guthrie's opinion that undue emphasis has been placed elsewhere is evidenced by his statement: "The literature of mental tests had, over twenty years ago, collected some ten thousand test titles and the number must be at least three times that figure by now" (4, p. 5).

In effect, educational psychologists have devoted disproportionately more effort to testing persons, than to basic learning experimentation.

In a recent review, Stroud indicates the need for a systematic behavior theory in education by stating: "Undoubtedly, one of the pressing needs in our field is a comprehensive learning theory (or behavior theory) that will encompass teaching and learning in school and embrace the broad phenomena of personality and adjustment. The educational psychologist is the logical person to attempt this. . . . A frank attempt to extend the current learning history theory of Hull, Tolman, or others, to learning in school is overdue" (13, p. 282).

Guthrie and Powers support this view with the following comment: "Research should increasingly deal with 'real' problems confronting educators.

---

*Received in the Editorial Office on March 10, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] A report of this research was presented, in part, at the 1953 meeting of the American Educational Research Association.

The authors wish to expressly thank both Dr. Nicholas Fattu, Director, Institute of Educational Research; and Dr. Wendell W. Wright, Dean, School of Education, Indiana University, for their support of basic experimentation.

Our continuing appreciation is extended, also, to Mr. Homer Terrell, principal of Central Elementary School, Martinsville, Indiana, who has not only granted permission for a series of classroom experiments, but, in addition, has materially aided in engendering a clearer understanding between psychologist and teacher, a trend we might add, that could profitably be perpetuated.

The writers are indebted, also, to the teachers, Miss Marlene Sullivan, who handled the experimental procedure for Class *A*, and Mrs. Jeanne Masters who replicated the operations in Class B.

The use of human subjects in investigations will not lend itself to the rigid control of all variables that is often possible in a simple animal study; nevertheless, results of such 'imperfect' observations should serve the practical function of pointing the way to more certain success in teaching" (5, p. 19).

Actually, animal experimentation and theory construction have made rapid progress during the last decade, presenting the educational psychologist with many concepts that appear to be testable under classroom conditions.

More specifically, the advantage in resistance to extinction of partial reinforcement over continuous reinforcement has received considerable attention from both experimenters and theorists since Skinner's 1938 book (12) and Humphrey's report (6) of the phenomenon in 1939. As Mowrer (9) points out, it was long assumed that the basic truth concerning learning was embodied in the law of exercise, which held that every performance of a habit necessarily strengthened that habit. But evidence has gradually accumulated which suggests that "exercise," as such, has little to do with the strength of a habit. It appears rather that the presence or absence of reinforcement is what is influential. The proposition that *every* rewarded repetition of a habit strengthens it is, however, open to question.

Numerous "partial" reinforcement experiments, however, offer an obstacle to the acceptance of this viewpoint. For example, in an earlier work Pavlov (10) describes an experiment performed in his laboratory in which conditioned salivation in a dog was established with food reinforcement on every other trial and on every third trial. Mowrer and Jones (8) compared extinction after acquisition with fixed ratios of 4/1 (four bar presses to one reinforcement), 3/1, 2.5/1, 2/1, and 1/1 (100 per cent reinforcement). Using 140 reinforcements in the training, they found large differences in responses in extinction, with the larger ratio associated with the greater number of responses when single bar presses were counted.

Generally speaking, acquisition may be quite efficient under various types of partial reinforcement, but a number of experimenters have reported results which indicate that extinction is slower following partial than following 100 per cent reinforcement. Jenkins and Stanley in their review of the literature on partial reniforcement make this statement:

> It is quite straightforward in view of the consistency of the data. All other things equal, resistance to extinction after partial reinforcement is greater than that after continuous reinforcement when behavior strength is measured in terms of single responses. A necessary condition for the operation of this principle appears to be massing of training trials or responses, at least with infra-human subjects (7, p. 222).

The action of an important variable in the study of partial reinforcement has been demonstrated by Sheffield (11). She found that resistance to extinction was slightly (not significantly) better after continuous reinforcement when the training trials had been spaced. With massed practice, the findings were reversed, and the typical difference favoring the partial group was obtained.

B. THE PROBLEM

The data presented in the Sheffield experiments appear to be relevant to a situation educators face in the classroom. Partial reinforcement is certainly more like the everyday life situation than 100 per cent reinforcement. It is rare, indeed, that every response leads to the presentation of a reinforcing stimulus. Continuous reinforcement, then, seems to be the exception, not the rule.

It was previously pointed out that the superiority of partial renforcement over continuous reinforcement with respect to greater resistance to extinction appears dependent upon whether the trials are massed. However, one is prone to question whether classroom "reinforcements" are meted out in a massed fashion. It is probably safer to assume, for the present, that, in general, classroom "reinforcements" are administered in somewhat of a spaced or distributed manner. It is almost certain that school room reinforcements are not dealt out in a massed fashion.

The question to be answered in the present experiment was: Would partial reinforcement have less advantage in resistance to extinction if classroom subjects were used with trials widely spaced?

The deduction to be made from Sheffield's experiment is where trials are widely spaced, no statistically significant differences will exist in extinction between groups given partial or continuous reinforcement. It is this deduction which the present set of experiments is intended to test.

C. A NOTE ON METHOD

It is well known that the results of an experiment can be evaluated only in the light of the measures which have been taken to insure operational isolation of the variables under consideration. The only completely satisfactory method of achieving such isolation is control of irrelevant variables. However psychologists have frequently been unable to secure complete control of experimental conditions and have resorted to the use of large numbers of subjects in the hope that the effects of uncontrolled factors would average out. In classroom situations, contrary to some opinions, it is often difficult to obtain a large sampling of subjects. In order, then, to investigate the phenomena under consideration, it was necessary to obtain a rather large number of "responses" from each subject over a relatively extended period of time. The general design, which will be described in detail later, is such that nearly every conclusion from the entire study is supported by the results of two independent experiments. Wherever feasible, statistical tests of significance have been applied. It has been assumed that only when the probability attaching to a value of F yielded by a group of measures is less than 0.01 can the probability that the observed differences are due solely to sampling errors be disregarded. Such a criterion of statistical significance is especially relevant to a program of classroom experimentation where the probability of experimental error is greatly increased.

D. GENERAL PROCEDURE

The subjects, third grade pupils at Central Elementary School in Martinsville, Indiana, were divided as indicated in Table 1 into three subgroups.

TABLE 1

SUMMARY OF EXPERIMENTAL PROCEDURES*

	Days		
Group	1	2-7	8-12
I	Pre-test measure for all *S*s.	Verbally reinforced each day.	Extinction (no verbal reinforcement for the task).
II		Verbally reinforced on days 2, 4, and 6.	
III		Never directly reinforced but in presence of the reinforcing stimuli for 11 days.	

*Two experiments were carried out independently of each other in separate classrooms, employing the operations described in the above table.

Day 1 was devoted to obtaining an initial measure of ability for the three groups in routine computational tasks. The assignment of subjects to the three groups was done randomly, using Tippett's tables of random sampling numbers.

Group I, designated as the continuous reinforcement group, consisted of one-third of a third grade class. This group was reinforced verbally on Days 2-7; that is, immediately before their regular teacher distributed the daily task, the names of the subjects in this group were called, and they were required to come to the front of the room and face the remainder of the class. The teacher then verbally reinforced them for the excellency of their work of the preceding day, as shown both in improvement made and general superiority over the other members of the class. Group II, as designated in

Table 1, received partial reinforcements which were given on Days 2, 4, and 6. Simply then, on Days 2, 4, and 6, Groups I and II were called to the front of the room to receive verbal reinforcement from their teacher, while on Days 3, 5, and 7, Group I alone, received the reinforcement.

Group III merely worked in the presence of the reinforcing stimuli, and was not directly reinforced by the teacher at any time during the experiment. Extinction, operationally defined for the purposes of the experiment as the withholding of verbal reinforcement for a specified task, started on Day 8 and continued for five days.

All subjects were required to participate in the task as part of the classroom routine, and, so far as can be determined, were naïve with respect to the fact that an experiment was being carried out.

The task was, in the main, a routine arithmetical one. It consisted primarily of one-digit number combinations with several two-digit numbers toward the latter sections of the work sheets.

Each subject was allowed five minutes a day to work the problems, there being enough problems, however, so that the most rapid subject was not able to finish before the time limit.

The task was administered at approximately the same time each day. The time between each replication was 24 hours, except for Days 5-6 and 10-11, 72 hours between these replications was necessary due to two week-end breaks.

Teachers were cautioned to avoid individual remedial work dealing with arithmetic. An attempt was also made to maintain in the experimental procedure some semblance of the normal classroom climate. For instance, the subjects were referred to by their "nick-names." The teacher avoided mentioning that they were participating in an experiment, and the task was integrated into the routine followed by that particular class.

The verbal reinforcement was similar to that used by Auble and Mech (1, 2) in previous classroom investigations. The teacher was required to adhere to the following verbal reinforcement with the appropriate subjects during the "acquisition" phase of the experiment:

> You children did very nice work on the arithmetic problems yesterday and are improving faster than the rest of the class. Let me see how much better you can do today. Try to avoid any careless mistakes, and try to do as many problems as you can.

The assumption, of course, is that the verbal stimuli from the teacher, designated as belonging to a class "reinforcing," are stimuli that will increase the probability that a particular class of response will occur more frequently than if these stimuli were withheld.

The general procedure described was followed in the execution of two experiments, hereafter referred to as *A* and *B*. Experiment *B* was an attempt to replicate Experiment *A*. Both experiments were carried out in separate third-grade classrooms, and, for practical purposes, should be considered as independent of each other.

E. Results and Discussion

For the purpose of maintaining continuity the results of the two experiments, *A* and *B*, will be presented separately. It should be recalled, however, that the operations in both experiments were intended to be identical.

1. *Experiment A*

Figure 1 shows the mean number of correct responses made by Groups I, II, and III during six days of acquisition and five days of extinction. From Figure 1 it is evident that the three groups made more correct responses during the six days of acquisition, when compared with their Day 1 per-

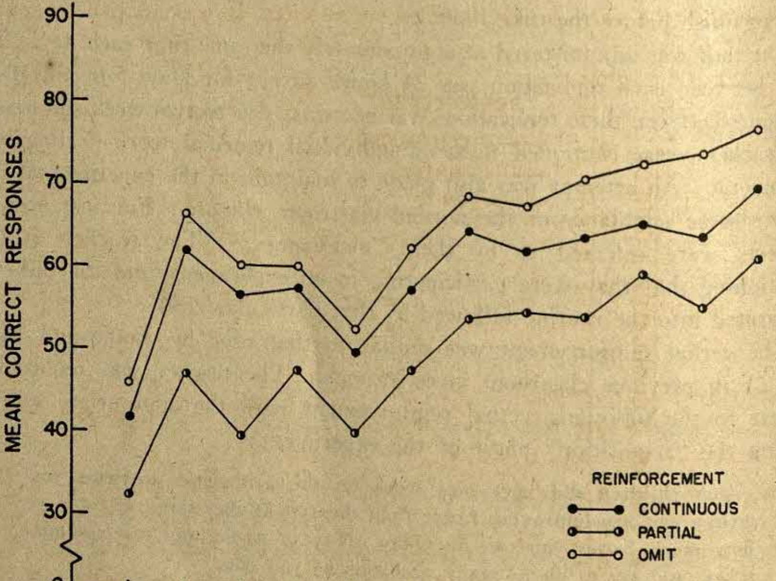


FIGURE 1

Mean Number of Correct Responses Made by Three Groups of *Ss* During Six Days of "Acquisition" and Five Days of "Extinction" (Group $N = 9$)

formance levels. Although there were no inversions in daily performance it is clear from the "slumps" on Days 3 and 5 that the groups fluctuated.

The groups, however, retained their original order of correct responding throughout the acquisition and extinction phases. It is interesting to note that the performance of Group II increased in mean number of correct responses on Days 2, 4, and 6, the only days when verbal reinforcement was administered to them.

During extinction the group curves appear more stable than in acquisition, with the groups still maintaining their relative ranks during this period.

It is clear that these data in their present form were not amenable to a straightforward statistical analysis, the reason being that the successive means and variances were proportional, and the distributions somewhat skewed. This occurs quite frequently when working with correct and error responses, and is to be expected. The use of an analysis of covariance at this point is not justifiable because of the possibility of violating several basic assumptions.

Bartlett (3) recommends a square root transformation when the distribution is such that the means and variances tend to be proportioned to one another within the various experimental conditions. It was decided to apply

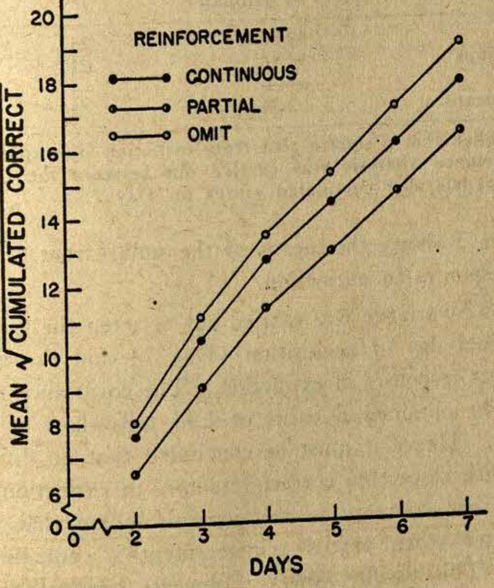


FIGURE 2
MEANS OF THE SQUARE ROOTS OF THE DAILY CUMULATED CORRECT RESPONSES FOR THREE GROUPS OF *S*s DURING SIX DAYS OF "ACQUISITION" (GROUP $N = 9$)

the appropriate square root transformation to all data in both experiments, so as to more readily meet the assumption underlying the model before applying covariance analyses.

Figure 2 shows the acquisition plot of the means of the square roots of the daily cumulated correct responses. The curves in Figure 2 appear to follow the theoretical prediction that no differences will exist between the continuous and partial reinforcement groups. It is also interesting to note that Group III, which was simply in the presence of the verbal reinforcements received by Groups I and II, maintained a rather consistent rate of correct responding during acquisition.

An analysis of covariance employing the correlation of transformed individual totals on the initial measure with the cumulated transformed totals of the same *S*s on the final day of acquisition indicated that the group differences were not significant. Table 2 shows the results of the covariance analysis. Figure 3 shows the means of the square root of the daily cumulated correct responses in extinction.

TABLE 2
EXPERIMENT *A*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN CORRECT RESPONSES DURING SIX DAYS OF ACQUISITION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	99.00	25	
Within groups	96.96	23	
Adjusted means	2.04	2	.24

*It will be recalled that a square root transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 1 scores and the transformed cumulated scores on Day 7.

An analysis of covariance was carried out between the transformed correct scores on the final day of acquisition (Day 7) and the total transformed cumulated correct responses in extinction. The covariance results are shown in Table 3. The obtained *F* value of 1.44 falls short of being significant at the .01 level. Hence it must be concluded that no differences exist between groups with respect to correct responses in extinction.

These results are contrary to the general finding that resistance to extinction is greater when partial reinforcement is compared to continuous reinforcement. This result, however, appears compatible with Sheffield's (11) finding that no differences will exist when the trials are spaced instead of massed.

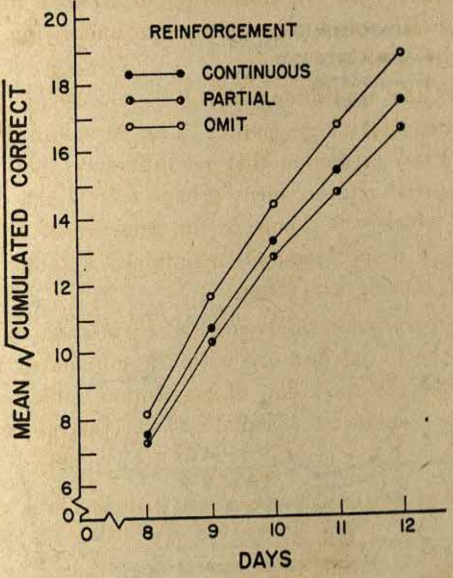


FIGURE 3

Means of the Square Roots of the Daily Cumulated Correct Responses for Three Groups of *S*s During Five Days of "Extinction" (Group $N = 9$)

TABLE 3

Experiment *A*: Analysis of Covariance for Differences in Correct Responses During Six Days of Acquisition*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	124.27	25	
Within groups	110.46	23	
Adjusted means	13.81	2	1.44

*A square root transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 7 acquisition responses and the transformed total cumulated responses in extinction.

Beginning the analysis for error response, Figure 4 shows the number of error responses made by the three groups during six days of acquisition and five days of extinction. From Figure 4 it is clear that many erratic reversals have taken place, especially during the acquisition phase.

In order to test for differences between groups with regard to error scores it was necessary to use a $\sqrt{X + .5}$ transformation. Bartlett (3) suggests this

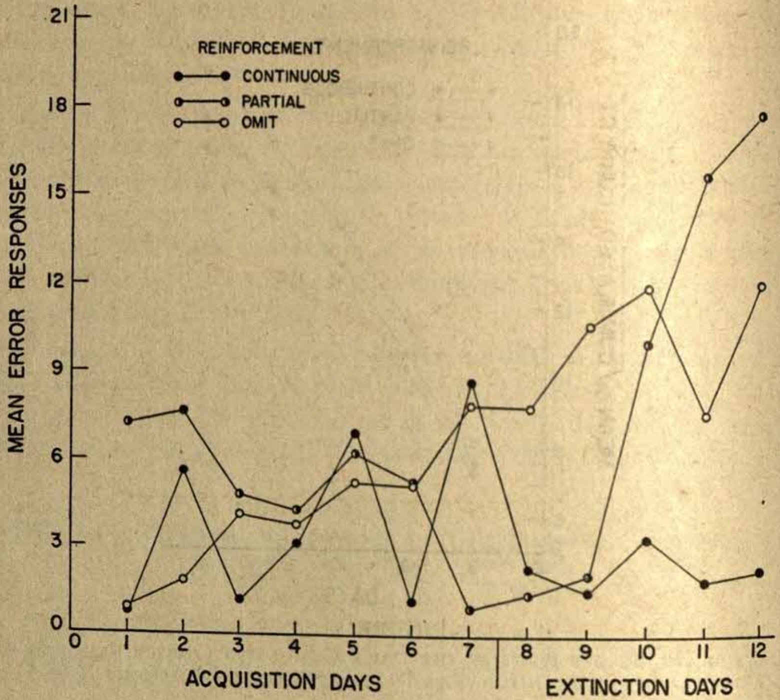


FIGURE 4
MEAN NUMBER OF ERROR RESPONSES MADE BY THREE GROUPS OF *S*s DURING SIX DAYS OF "ACQUISITION" AND FIVE DAYS OF "EXTINCTION" (GROUP $N = 9$)

transformation when many of the scores are zero. Figure 5 shows the transformed cumulated error responses for six days of acquisition.

Table 4 summarizes the results of the covariance analysis of the transformed error scores for the six days of acquisition.

TABLE 4
EXPERIMENT *A*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN ERROR RESPONSES DURING SIX DAYS OF ACQUISITION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	407.88	25	
Within groups	389.84	23	
Adjusted means	18.04	2	.53

*$\sqrt{X + .5}$ transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 1 error scores and the transformed cumulated error scores on Day 7.

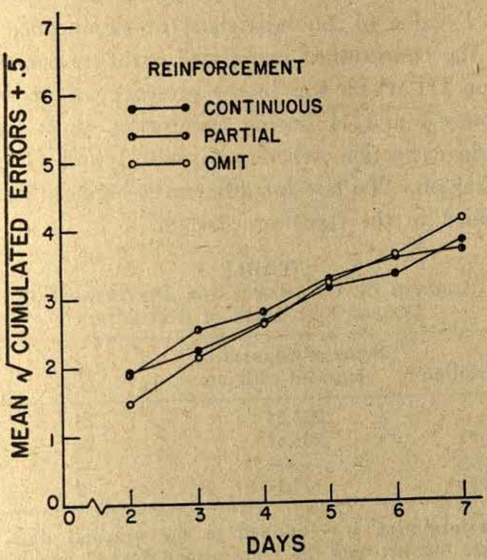


FIGURE 5

MEANS OF THE $\sqrt{X + .5}$ TRANSFORMED DAILY CUMULATED ERROR RESPONSES FOR THREE GROUPS OF *S*S DURING SIX DAYS OF "ACQUISITION" (GROUP $N = 9$)

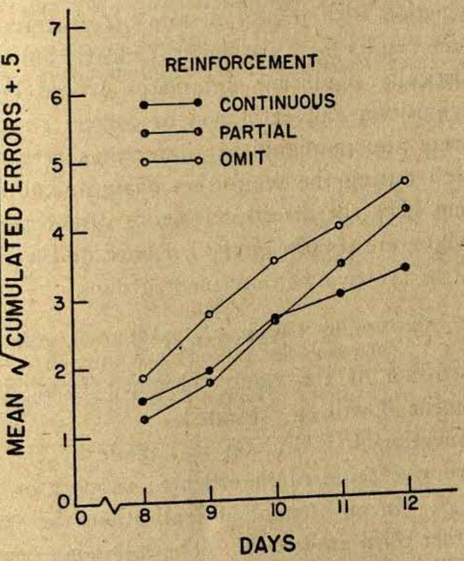


FIGURE 6

MEANS OF THE $\sqrt{X + .5}$ TRANSFORMED DAILY CUMULATED ERROR RESPONSES FOR THREE GROUPS OF *S*S DURING FIVE DAYS OF "EXTINCTION" (GROUP $N = 9$)

The obtained F value of .66 falls short of significance at the .01 level. Figure 6 shows the transformed cumulated error responses for five extinction days. Group III which was in the presence of the reinforcing stimuli received by Groups I and II during acquisition, gives some indication of increased errors in extinction, whereas Groups I and II cross over on the third day of extinction. To test for differences between groups a covariance analysis was applied to the transformed data.

TABLE 5

EXPERIMENT *A*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN ERROR RESPONSES DURING FIVE DAYS OF EXTINCTION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	250.25	25	
Within groups	236.68	23	
Adjusted means	13.57	2	.66

*$\sqrt{X + .5}$ transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 7 acquisition responses and the transformed total cumulated responses in extinction.

The obtained F value of .52 does not approach statistical significance at the .01 level, hence it must be concluded that no differences existed between the groups in extinction with respect to rate of error responses.

A summary of the results for Experiment *A* clearly indicates the following:

(*a*). No statistically significant differences existed between the groups with regard to acquisition and extinction of correct responses.

(*b*). There were also no significant differences between the groups for error responses both during the acquisition phase and during extinction.

(*c*). *S*s (Group III) not directly receiving verbal reinforcement during acquisition, but who were simply in the presence of the reinforcing stimuli, performed as well as the two reinforcement groups.

2. *Experiment B*

Prior to a discussion of the results obtained from Experiment *A*, the data from Experiment *B* will be presented.

As previously mentioned, a decision was made by the investigators that additional evidence was required in order to support or refute the results from Experiment *A*. In an attempt at replication, the same procedure was carried out in another third grade class. The following results were obtained.

Figure 7 shows the composite mean curves for acquisition and extinction of correct responses for the 12-day period.

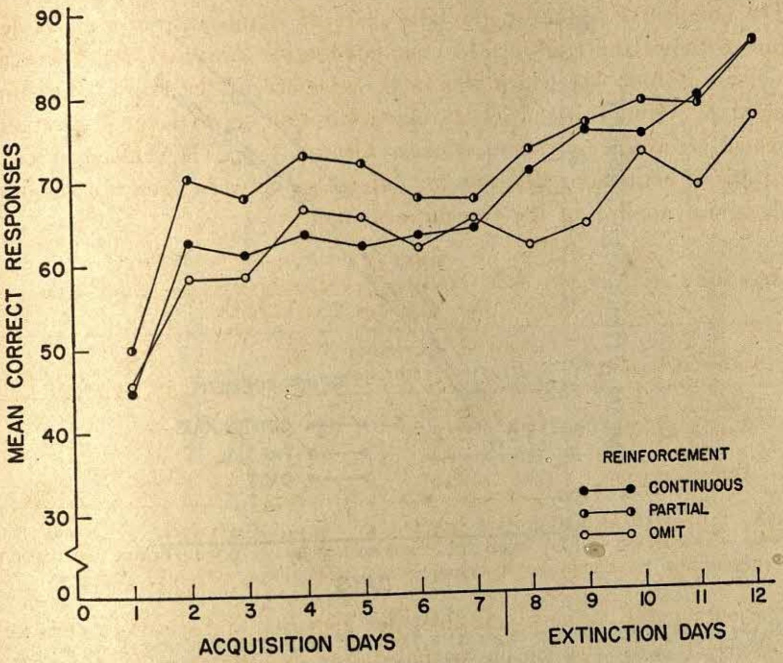


FIGURE 7
MEAN NUMBER OF CORRECT RESPONSES MADE BY THREE GROUPS OF *S*s DURING SIX DAYS OF "ACQUISITION" AND FIVE DAYS OF "EXTINCTION" (GROUP N = ACQUISITION 8; EXTINCTION 6)

However, prior to justifiably applying a statistical analysis a square root transformation was employed to reduce heterogeneity of variance and to correct somewhat for the skewed distribution. Figure 8 shows the acquisition plot of the means of the square roots of the cumulated correct responses. The covariance analysis for differences is shown in Table 6.

TABLE 6
EXPERIMENT *B*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN CORRECT RESPONSES DURING SIX DAYS OF ACQUISITION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	46.27	22	
Within groups	45.84	20	
Adjusted mean	.43	2	.09

*A square root transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 1 error scores and the transformed total cumulated correct responses on Day 7.

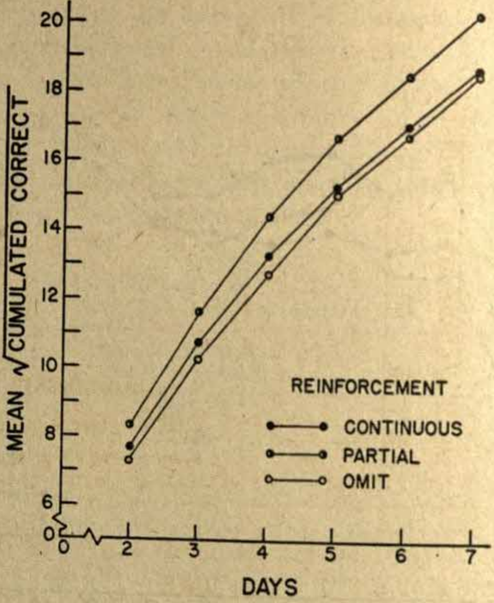


FIGURE 8
MEANS OF THE SQUARE ROOTS OF THE DAILY CUMULATED CORRECT RESPONSES FOR THREE GROUPS OF *Ss* DURING SIX DAYS OF "ACQUISITION" (GROUP $N = 8$)

The *F* value of .09 for 2 and 20 degrees of freedom is not statistically significant, and it is concluded that the groups did not differ with respect to correct rate of responding during the acquisition phase.

Figure 9 shows the means of the transformed cumulated correct responses during five days of extinction. It should be pointed out that the extinction curves were based on six subjects in each group instead of the eight reported during acquisition. Two subjects in each group were absent during the extinction days. Table 7 presents the results of the covariance analysis for the mean transformed cumulated correct responses.

TABLE 7
EXPERIMENT *B*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN CORRECT RESPONSES DURING FIVE DAYS OF EXTINCTION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	27.69	16	
Within groups	20.30	14	
Adjusted means	7.39	2	2.55

*A square root transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 7 acquisition responses and the transformed total cumulated responses in extinction.

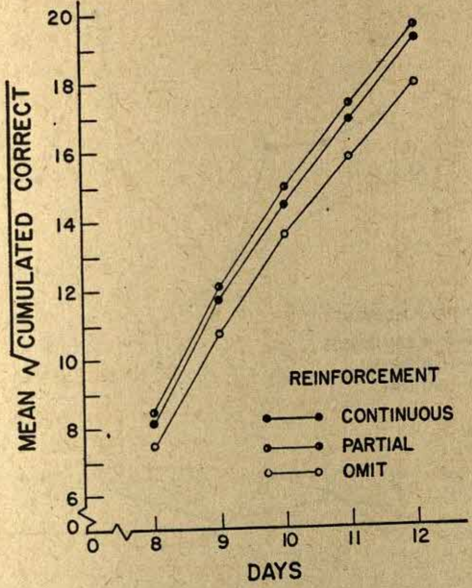


FIGURE 9

Means of the Square Roots of the Daily Cumulated Correct Responses for Three Groups of *S*s During Five Days of "Extinction" (Group $N = 6$)

The obtained *F* value of 2.55 for 2 and 14 degrees of freedom falls short of being statistically significant at the .01 level, and it is concluded that no differences exist between groups for correct responses in extinction.

Figure 10 shows the performance curves for acquisition and extinction of error responses over a 12-day period. As with the previous error curves these, also, present a problem with respect to reducing the heterogeneity of variances prior to testing for statistical differences. The previously referred to $\sqrt{X + .5}$ transformation was used. The means of the transformed cumulated scores are shown in Figure 11.

Although the heterogeneity of variance was markedly reduced as a result of the transformation, it was apparent that the variances were not sufficiently homogeneous to permit full confidence in the interpretation.

Table 8 shows the results of the covariance analysis for the transformed error responses during six days of acquisition.

Although the obtained *F* ratio of .18 was clearly not significant the value was probably somewhat higher than if the homogeneity of variance assumption underlying the model was completely satisfied.

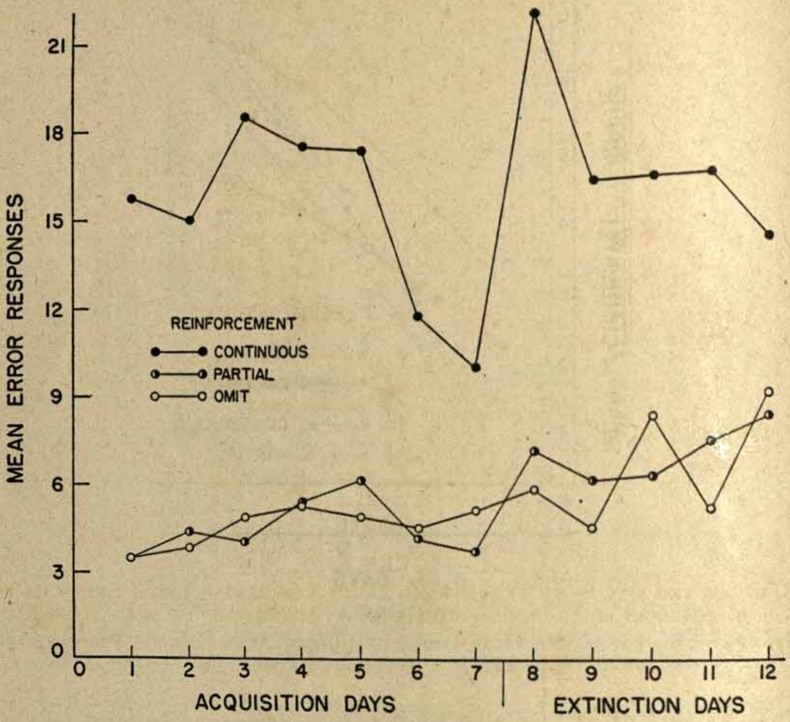


FIGURE 10

MEAN NUMBER OF ERROR RESPONSES MADE BY THREE GROUPS OF *Ss* DURING SIX DAYS OF "ACQUISITION" AND FIVE DAYS OF "EXTINCTION" (GROUP N = ACQUISITION 8; EXTINCTION 6)

TABLE 8

EXPERIMENT *B*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN ERROR RESPONSES DURING SIX DAYS OF ACQUISITION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	69.97	22	
Within groups	68.71	20	
Adjusted means	1.26	2	.18

*$\sqrt{X + .5}$ transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 1 error scores and the transformed cumulated error scores on Day 7.

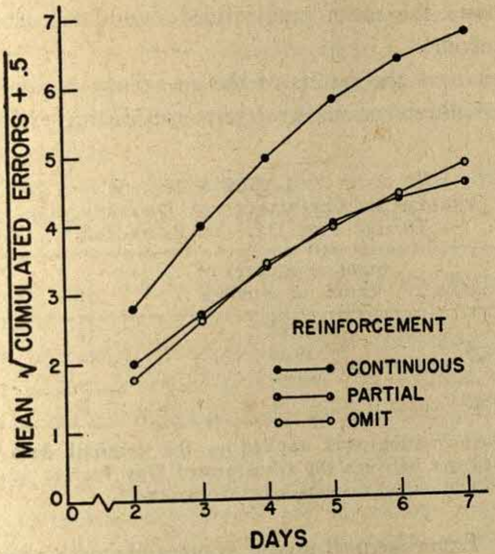


FIGURE 11

MEANS OF THE $\sqrt{X + .5}$ TRANSFORMED DAILY CUMULATED ERROR RESPONSES FOR THREE GROUPS OF *S*s DURING SIX DAYS OF "ACQUISITION" (GROUP $N = 8$)

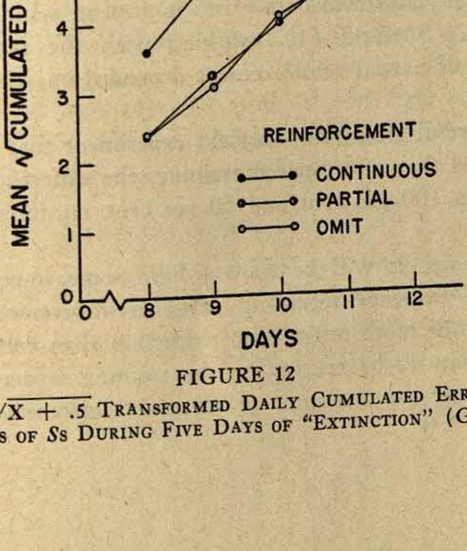


FIGURE 12

MEANS OF THE $\sqrt{X + .5}$ TRANSFORMED DAILY CUMULATED ERROR RESPONSES FOR THREE GROUPS OF *S*s DURING FIVE DAYS OF "EXTINCTION" (GROUP $N = 6$)

Figure 12 shows the mean transformed cumulated error scores during five days of extinction.

Table 9 summarizes the results of the covariance analysis. The obtained *F* ratio of .01 for differences in error responses during extinction was clearly not significant.

TABLE 9
EXPERIMENT *B*: ANALYSIS OF COVARIANCE FOR DIFFERENCES IN ERROR RESPONSES DURING FIVE DAYS OF EXTINCTION*

Source of variation	Sums of squares of errors of estimate	*df*	*F*
Total	10.86	16	
Within groups	10.85	14	
Adjusted means	.01	2	.01

*$\sqrt{X + .5}$ transformation was applied to the original data. The covariance analysis was carried out between the transformed Day 7 error scores and the transformed total cumulated error responses in extinction.

The results of Experiment *B* appear compatible with those obtained from Experiment *A*. A main finding was that after spaced training, the difference in resistance to extinction between the continuous and partial reinforcement groups was not significant. Eight covariance analyses were carried out for error and correct responses during acquisition and extinction, no statistically significant differences were apparent.

As stated previously Experiments *A* and *B* were based on a hypothesis that would provide a test for a stimulus response explanation for the greater resistance to extinction that has been found after practice, with 50 per cent reinforcement as compared with 100 per cent reinforcement.

Unfortunately no classroom data exist, at present, with which to compare the cited results. Sheffield (1), working with the rat, found that the differential effect of partial reinforcement depended on whether practice was massed or spaced.

An important result from the Sheffield experiment that is relevant to the present paper, was that after spaced training, the difference in resistance to extinction between 100 per cent and 50 per cent reinforcement groups was not significant.

Sheffield's findings, it will be shown, have some import for educators. The experimental evidence regarding partial reinforcement has quite generally shown it to be more resistant to extinction than continuous reinforcement. This appears to be true when the training is massed and not distributed. However, the analogue of "massed reinforcement" is not often encountered in an everyday classroom setting. Moreover, when reinforcements are given they are apt to be administered in a spaced pattern rather than massed. As pointed out in the previous section, the test of this interpretation consisted of "controlling" some of the after effects of reinforcement by the spacing of trials.

The assumption was that if classroom subjects were used with trials widely spaced most of the effects of reinforcement would have "dissipated" by the start of the next trial, theoretically making the stimuli much the same for the reinforcement groups. This means that the group receiving reinforcement on each day and the group receiving partial reinforcement would be under more similar patterns of stimuli than when trials were massed. Similarly the change in conditions when "extinction" was started would be more nearly the same for the two groups; and consequently, the difference in resistance to extinction should be reduced. Therefore it follows that the extent to which pattern of reinforcement enters into the stimulus pattern on a given trial will be minimized if trials are spaced far apart. The wide spacing of trials, then, should obliterate any advantage partial reinforcement might have.

The results of both experiments, *A* and *B*, demonstrate that partial reinforcement was not superior to continuous reinforcement with regard to differential responding during extinction. This result appears attributable to the widely spaced trials used. An additional interesting result was that the group simply in the presence of the reinforcing stimuli never directly receiving it, performed as well, both during acquisition and extinction, as did the two reinforcement groups.

In a word, there were no statistically significant differences found as evidenced by the eight covariance analyses, for the two experiments reported. The groups were not different either during acquisition or extinction for both error and correct responses.

While similar results were obtained from two separate experiments they should be interpreted, for the present time, as being provocative of an intensive program of classroom learning experimentation.

F. SUMMARY AND CONCLUSIONS

The present study began with the belief that progress could be made toward evolving a behavior theory for education, by virtue of testing certain concepts derived from learning theory. Specifically, the question to be answered in the present experiment was: Would partial reinforcement have less advantage in resistance to extinction if classroom subjects were used with trials widely spaced? The deduction to be made from existing experimental literature on the infra-human level is where trials are widely spaced, no statistically significant differences will exist in extinction between groups given partial or continuous reinforcement.

The study of partial reinforcement, which is certainly more like the every day life situation than 100 per cent reinforcement, clearly merits attention in the classroom situation. Generally speaking, rarely if ever does every response lead to the presentation of reinforcing stimulus. Continuous reinforcement, then, seems to be the exception, not the rule.

As previously pointed out, the superiority of partial reinforcement over continuous reinforcement with respect to greater resistance to extinction appears to be a function of whether the trials are massed or spaced. The variable of spaced reinforcements is indeed an important one if generalizations are to be made regarding classroom situations. It is more tenable to assume that when reinforcements are administered in the school, they are given in a fashion that leans toward a spaced rather than massed pattern. However, it is almost certain that schoolroom reinforcements are not dealt out in a massed fashion. Therefore, it is risky to conclude that partial reinforcement will prove more resistant to extinction in a school situation because of the spacing factor.

To test whether partial reinforcement would have less advantage in resistance to extinction in a classroom situation when the trials were widely spaced, the following procedure was used: A class of third grade pupils at Central Elementary School in Martinsville, Indiana, was divided into three groups. Group I, designated as the continuous reinforcement group consisted of one-third of a third-grade class. This group was verbally reinforced on Days 2-7, that is, immediately before their regular teacher distributed the daily task, the names of the subjects in this group were called, and they were required to come to the front of the room and face the remainder of the class. The teacher then verbally reinforced them for the excellency of their work of the preceding day, as shown both in improvement made and general superiority over the other members of the class. Group II, belonging to the partial reinforcement group, was reinforced on Days 2, 4, and 6. Simply then, on Days 2, 4, and 6, both Groups I and II were called to the front of the room to receive verbal reinforcement from the teacher, while on Days 3, 5, and 7, Group I alone received reinforcement. Group III worked in the presence of the reinforcing stimuli, and was not directly reinforced by the teacher at any time during the experiment.

Day 1 was devoted to obtaining an initial measure of ability for the three groups in a routine computational task. The assignment of subjects to the three groups was done from a table of random sampling numbers.

Extinction, operationally defined for the purposes of the experiment as the withholding of verbal reinforcement for a specified task, started on Day 8 and continued for five days.

All subjects were required to participate in the task as part of the classroom routine, and, so far as can be determined, were naïve with respect to the fact that an experiment was being carried out.

The task was, in the main, a routine arithmetical one, consisting primarily of one-digit number combinations with several two-digit numbers toward the latter sections of the work sheets. Each subject was allowed five minutes a day to work the problems, there being enough problems, however, so that the most rapid subject was not able to finish before the time limit.

The task was administered at approximately the same time each day. The time between each replication was 24 hours, except for 5-6 and 10-11, 72 hours between these replications was necessary due to two week-end breaks.

The general procedure described was followed in the execution of two experiments. Experiment *B* was an attempt to replicate Experiment *A*. Both were carried out in separate third-grade classrooms.

The main conclusions supported by the results of two separate experiments were that for the population studied:

a. The difference in resistance to extinction after spaced training between continuous and partial reinforcement groups was not significant for either error or correct responses.

b. There were no significant differences in level of performance on the acquisition trials for correct or error responses.

c. A mediation of the reinforcing stimulus appeared to exist for the subjects who did not directly receive it, but who were in the presence of the reinforcement.

REFERENCES

1. AUBLE, J. D., & MECH, E. V. Quantitative studies of verbal reinforcement in classroom situations: I. Differential reinforcement related to the frequency of error and correct responses. *J. of Psychol.*, 1953, **35**, 307-312.
2. AUBLE, J. D., & MECH, E. V. Quantitative studies of verbal reinforcement in classroom situations: II. Occurrence of error and correct responses related to a change in stimulus conditions. *J. Genet. Psychol.*, (in press).
3. BARTLETT, M. S. The use of transformations. *Biometrics*, 1947, **3**, 39-52.
4. GUTHRIE, E. R. Psychological facts and psychological theory. *Psychol. Bull.*, 1946, **43**, 1-20.
5. GUTHRIE, E. R., & POWERS, F. F. Educational Psychology. New York: Ronald Press, 1950.

6. HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. Exp. Psychol.*, 1939, **25**, 141-158.

7. JENKINS, W. O., & STANLEY, J. C. Partial reinforcement. A review and critique. *Psychol. Bull.*, 1950, **47**, 193-234.

8. MOWRER, O. H., & JONES, H. M. Habit strength as a function of the pattern of reinforcement. *J. Exp. Psychol.*, 1945, **35**, 293-311.

9. MOWRER, O. H. Learning Theory and Personality Dynamics. New York: Ronald Press, 1950.

10. PAVLOV, I. P. Conditioned Reflexes. (Trans. by G. V. Anrep.) London: Oxford Univ. Press, 1927.

11. SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. Exp. Psychol.*, 1949, **39**, 511-525.

12. SKINNER, B. F. The Behavior of Organisms. New York: Appleton-Century, 1938.

13. STROUD, J. B. Educational psychology. In C. P. Stone (*Ed.*), *Annual Review of Psychology.* Stanford, Calif.: Annual Reviews, Inc., 1950.

Institute of Educational Research
Indiana University
Bloomington, Indiana

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 187-194.

THE EFFECTS OF "SET" ON GROUP PERFORMANCE UNDER STRONG AUDITORY STIMULI*

Institute of Educational Research, Indiana University

E. VICTOR MECH, ROBERT W. SCHAERER, AND DONAVON AUBLE[1]

A. THE PROBLEM

The emphasis which is given to noise abatement campaigns as indicated by Ryan (5) implies a common assumption that noise is a barrier to efficient work. This interest in controlling noise, unfortunately, has far outdistanced the empirical evidence available. It is defensible to state, then, that both industry and education lack sufficient data relative to this problem which will provide a basis for generalizations.

Berrien (2), in a cogent review on the effects of noise, points out that the factors of motivation and suggestion merit more experimental attention than they have generally received.

Previous investigations by Mech (4) and Baker (1) demonstrate that the variable "set" appears to determine, in part, a person's performance under noise conditions. However, these experiments used individual *S*s and a noise level of but 70 db. intensity. It should be pointed out that the intensity level of 70 db. exceeds the average conversational tone by approximately 12 decibels.

The experiment being reported was concerned with two facets: (*a*) whether systematic group performance changes are possible in the presence of strong auditory stimuli as a function of a verbal "set"; (*b*) to obtain information with regard to whether noise stimuli at an accelerated intensity level exerts a deleterious effect on overt performance.

B. PROCEDURE

Three groups of 13 *S*s each were required to perform a coding task for 20 minutes each day. Each group worked five days under two conditions each day; one a quiet condition, the other a noise condition. Each condi-

---

*Received in the Editorial Office on March 20, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The authors wish to express their gratitude to both Dr. Wendell W. Wright, Dean, School of Education; and Dr. Nicholas Fattu, Director, Institute of Educational Research, Indiana University, for their continuing support of basic experimentation.

tion on each day consisted of a 10-minute work period. Thus each group served a total of 1 2/3 work hours in the experiment over the five-day period. The coding task consisted of pairing a set of numbers with given symbols. The number of correct responses during each 10-minute period served as the basic response measure.

The verbal noise used in the experiment was Volume I: "I Can Hear It Now," side No. 2. This record contains verbal narrations of selected news events between the 1933-1945 era. The record was played at an average intensity of 85 decibels, as determined by a sound level meter, Model 759-B. The measurement of 85 db. was made at the middlemost position at which the groups were required to work. Senior high school students from three physics and chemistry sections acted as subjects and were assigned randomly to the experimental groups. There was a 24-hour interval between all replications the noise and quiet conditions being alternated in a simple *ABBA* fashion.

Table 1 summarizes the experimental procedures for the three groups.

TABLE 1
SUMMARY OF EXPERIMENTAL PROCEDURES (GROUP $N = 13$)

Group	Day 1	Days 2-5
I	Pre-test measures for all *S*s in a coding task	Told only that they were serving in an experiment on the effects of noise.
II		*S*s were given a two-minute, tape-recorded "set" to perform better under quiet conditions.
III		*S*s were given two-minute, tape-recorded "set" to perform better under the noise condition.

Group I was told only that they were participating in an experiment on the effects of noise. On Days 2-5 prior to beginning the task, Group II was given a two-minute, tape-recorded verbal "set" to the effect that they were expected to work better under quiet conditions, previous studies showing that other high school students did better under quiet conditions.

Group III was also given a two-minute, tape-recorded verbal "set" prior to the task on Days 2-5. However, the instructions stated that this group was expected to work better under the noise conditions, that with a previous class noise was found to engender performance.

C. Results and Discussion

Figures 1 and 2 show the plots of the mean correct responses under quiet and noise for the three groups.

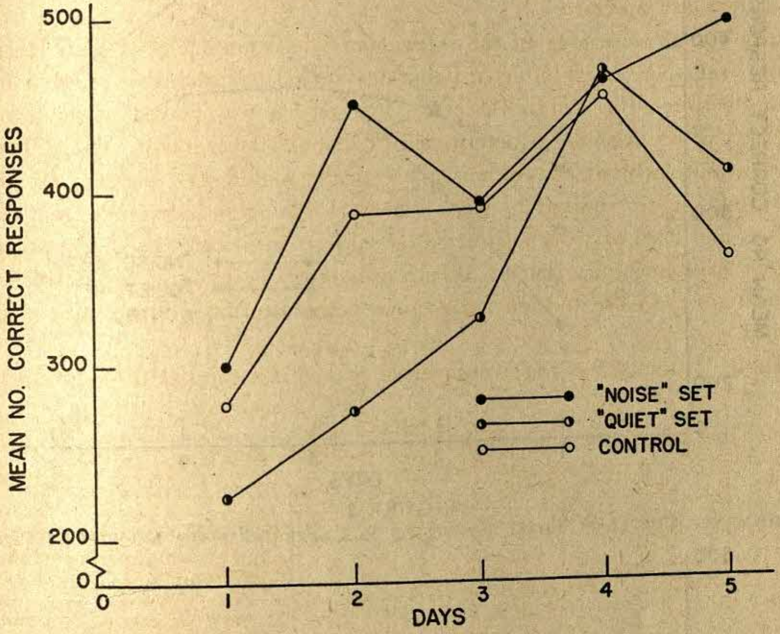


FIGURE 1
Performance Curves of Three Groups of *S*s Under Quiet for a Five-Day Period

Table 2 summarizes the covariance *F* values for Days 1 versus 5. The "sets" were not introduced until Day 2 and the covariance computations were made against the pre-test measure, Day 1.

Both covariance *F* values did not approach significance at the .05 level.

The next operation was to consider the intra-group differences with regard to performance under noise and quiet conditions. Figures 3, 4, and 5 show the performance curves of the three respective groups, i.e., the same *S*s working under both conditions.

TABLE 2
Covariance *F* Values for the Three Groups, Day 1 versus Day 5*

Treatment	Quiet	Noise
df	2,35	2,35
F	1.70	2.12

*The *F* values did not prove significant at the .05 level of confidence.

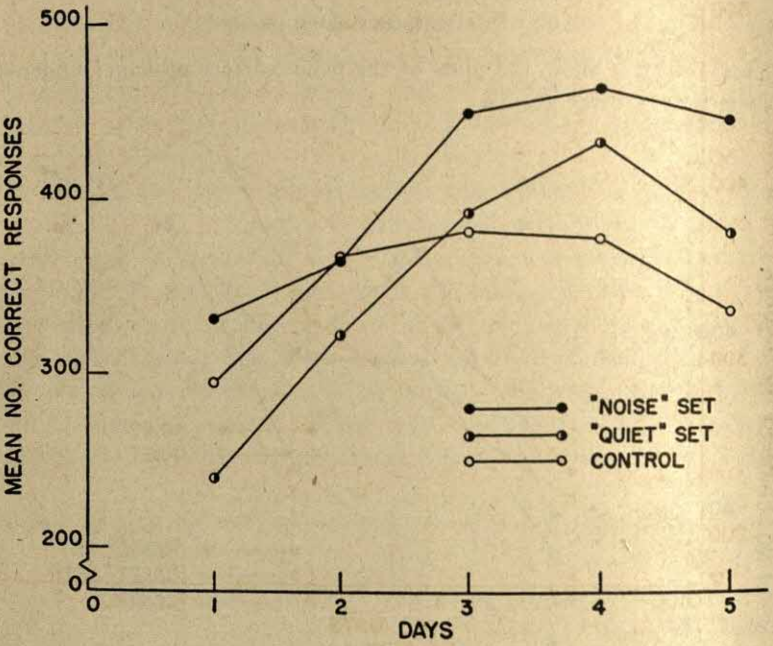


FIGURE 2
PERFORMANCE CURVES OF THREE GROUPS OF *S*s UNDER NOISE FOR A FIVE-DAY PERIOD

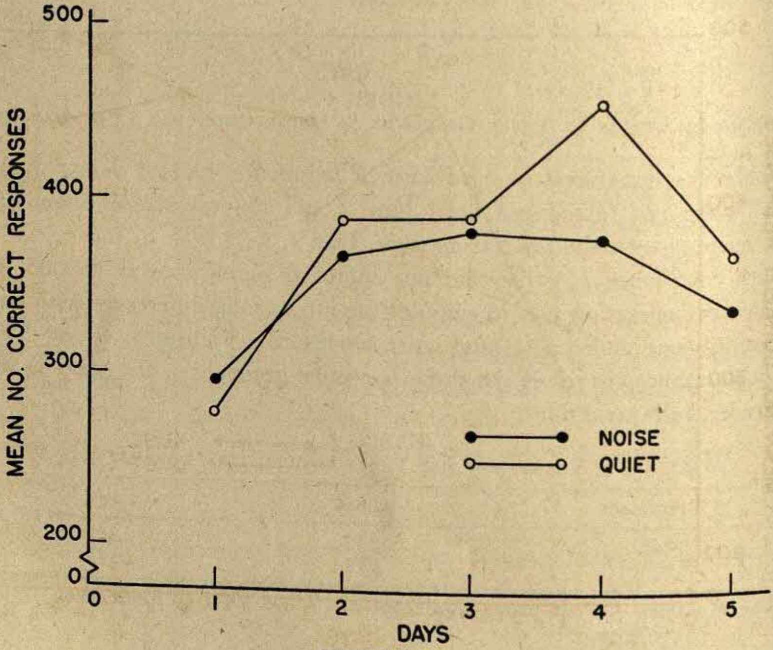


FIGURE 3
PERFORMANCE CURVES OF GROUP I: SAME *S*s WORKING UNDER NOISE AND QUIET EACH DAY

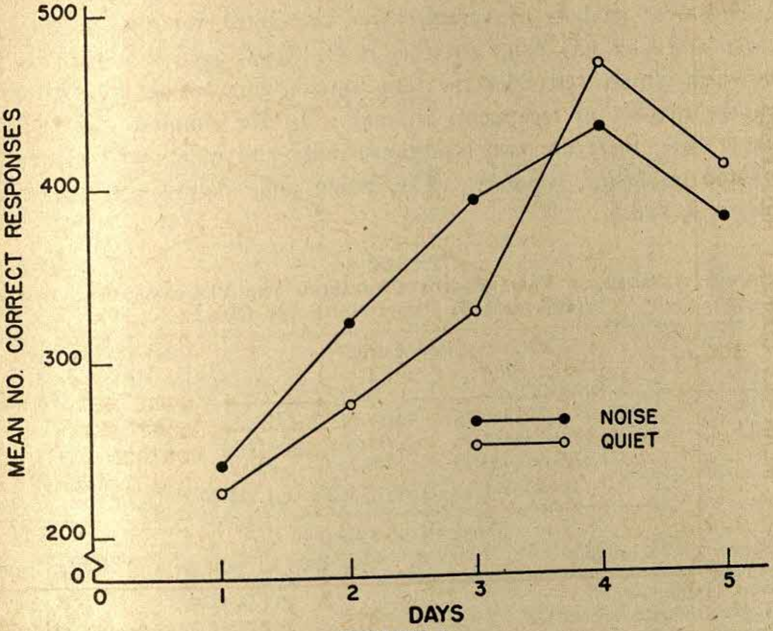


FIGURE 4
PERFORMANCE CURVES OF GROUP II: SAME *S*s WORKING UNDER NOISE AND QUIET EACH DAY

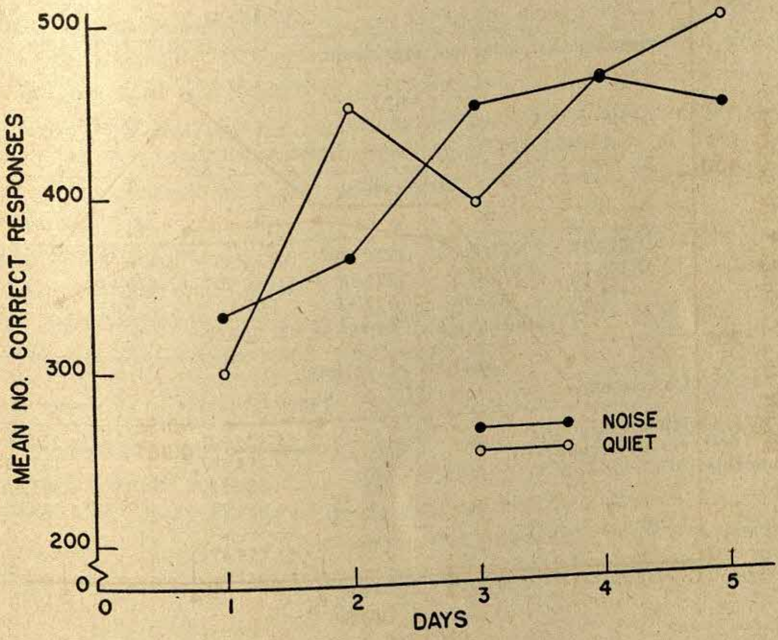


FIGURE 5
PERFORMANCE CURVES OF GROUP III: SAME *S*s WORKING UNDER NOISE AND QUIET EACH DAY

A cross-over analysis of variance was calculated for the three groups. Cochran and Cox (3) point out that in biological assay a design has been used which closely resembles the latin square, but has several advantages when the number of treatments is small. In the simplest case, which is apparent here, there are two conditions, noise and quiet, which have been alternated in *ABBA* sequence. The design and analysis are presented in Tables 3, 4, and 5.

TABLE 3
CROSS-OVER ANALYSIS OF VARIANCE FOR COMPARING THE PERFORMANCE OF GROUP I (CONTROL) *S*s UNDER NOISE AND QUIET

	Replications					
Rows	1	2	3	4	5	Totals
First	Q276.57	N366.86	Q389.71	N376.86	Q366.00	1,776.00
Second	N295.00	Q387.29	N381.00	Q453.43	N335.14	1,851.86
Totals	571.57	754.15	770.71	830.29	701.14	3,627.86

Treatment totals: N1,754.86, Q1,873.00

Analysis of variance

Source of variation	*df*	Sum of squares	Mean square
Columns (replications)	4	19,066.77	476.67
Rows (first versus second)	1	575.47	575.47
Treatments	1	1,395.71	1,395.71
Error	3	1,852.93	617.64
Total	9	22,890.88	

F-ratio for Treatments = 2.26; not significant.

TABLE 4
CROSS-OVER ANALYSIS OF VARIANCE FOR COMPARING THE PERFORMANCE OF GROUP II (QUIET "SET") *S*s UNDER NOISE AND QUIET

	Replications					
Rows	1	2	3	4	5	Totals
First	Q224.29	N321.86	Q326.57	N432.00	Q410.00	1,714.72
Second	N240.43	Q273.57	N391.86	Q468.57	N379.43	1,753.86
Totals	464.72	595.43	718.43	900.57	789.43	3,468.58

Treatment totals: N1,765.58, Q1,703.00

Analysis of variance

Source of variation	*df*	Sum of squares	Mean square
Columns (replications)	4	57,329.92	14,332.48
Rows (first versus second)	1	153.19	153.19
Treatments	1	391.62	391.62
Error	3	4,018.74	1,339.58
Total	9	61,893.47	

F-ratio for Treatments = 0.29; not significant.

TABLE 5

CROSS-OVER ANALYSIS OF VARIANCE FOR COMPARING THE PERFORMANCE OF GROUP III (NOISE "SET") *S*s UNDER NOISE AND QUIET

	Replications					
Rows	1	2	3	4	5	Totals
First	Q300.00	N364.14	Q393.86	N462.43	N444.43	1,964.86
Second	N331.86	Q450.00	N448.71	Q462.86	Q497.00	2,190.43
Totals	631.86	814.14	842.57	925.29	941.43	4,155.29

Treatment totals: N2,051.57, Q2,103.72

Analysis of variance

Source of variation	*df*	Sum of squares	Mean square
Columns (replications)	4	30,580.12	7,645.03
Rows (first versus second)	1	3,088.18	5,088.18
Treatments	1	271.96	271.96
Error	3	1,719.51	573.17
Total	9	37,659.77	

F-ratio for Treatments = 0.47; not significant.

To summarize the results: (*a*) no significant inter-group differences were evident for both noise and quiet conditions; (*b*) in comparing the intra-group trends, no group responded differentially in either direction.

The results indicate quite definitely that systematic *group* performance changes were not produced as a function of a verbal "set."

However this finding points out the difficulty encountered in controlling such a variable as "set" in a *group* situation. Previous investigations (1, 4) have demonstrated that systematic performances are possible under noise as a function of set, when individual *S*s were used, under noise levels of approximately 70 decibels. Unfortunately, no systematic data exist that deal with attempts to influence overt performance of *groups* under strong auditory stimulation.

Although the influence of "set" was ineffective on group performance at the relatively high noise level of 85 db., it is clear that noise stimulation of such accelerated intensity does not appear to have any necessary effect upon the execution of a routine task.

D. SUMMARY

The purpose of the experiment was to study the effects of "set" upon group performance under noise. The following procedure was used.

Three high school groups of 13 *S*s each worked at a routine coding task 20 minutes a day for five days under the following conditions. For 10 minutes of each session the groups worked under noise stimulation of 85 decibels, while for the remaining 10 minutes the groups worked under quiet conditions. Noise and quiet conditions were alternated in a simple *ABBA* sequence. One group was told only that they were serving in an experiment on the effects of noise. Of the remaining two groups, one was given a "set" to perform better under noise while the other was given a "set" to perform better under quiet.

Analysis of the data indicated: (*a*) The pre-task "sets" did not produce differential group results in the desired direction. (*b*) Verbal noise, of an accelerated intensity level does not appear to have any necessary effects upon routine performance.

REFERENCES

1. BAKER, K. H. Pre-experimental set in distraction experiments. *J. Gen. Psychol.,* 1937, **16**, 471-488.
2. BERRIEN, F. K. Relation of noise to the habitability of submarines in *Human Factors in Undersea Warfare.* Washington, D. C.: National Research Council, 1949 (Chap. 16).
3. COCHRAN, W. G., & COX, G. M. Experimental Designs. New York: Wiley, 1950. Pp. ix+454.
4. MECH, E. V. Factors influencing routine performance under noise: I. The influence of "set." *J. of Psychol.,* 1953, **35**, 283-298.
5. RYAN, T. A. Work and Effort: The Psychology of Production. New York: Ronald Press, 1947. Pp. vii+232.

Institute of Educational Research
Indiana University
Bloomington, Indiana

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 195-200.

A STUDY OF THE LEARNING CHARACTERISTICS OF PUBLIC HEALTH NURSES IN RELATION TO MENTAL HEALTH EDUCATION*: A PRELIMINARY REPORT

Westchester County Department of Health and Long Island University

RUTH CUMINGS AND MILTON V. KLINE

The Health Department of Westchester County early recognized and accepted responsibility for community mental health, and the possibilities inherent in integrating mental health concepts into its progressive and expanding public health nursing service. Because of this new trend in nursing service, and the problems encountered here (1) and elsewhere in such programs, it seems pertinent to examine the nature of mental health consultation and the methods and techniques that facilitate in-service education in mental health. It also seems necessary to understand the factors and conditions that influence public health nurses' learning and integration of such knowledge into their professional rôle and activities. As a part of a long range investigation of the rôle of the public health nurse in mental health education and practice, it seemed necessary to obtain the nurses' own evaluation of the program in which they had been involved for two years.

The specific purpose of this study has been to obtain a critical evaluation of an in-service mental hygiene consultation program by the nurse participants. This was undertaken through an interview analysis technique which was designed to measure the nature of the nurses' reaction to the program as a group, and in relation to empirical classifications of nurses according to their age, experience, marital status, education, and consultation relationships. From the variability of nurse response to the program, it was hoped that it might be possible to isolate clusters of nursing characteristics for further investigation in relation to their adaptation in mental health areas.

Each participating nurse who had been involved in the in-service mental hygiene program was seen in an interview situation designed to obtain her evaluation of the program and her attitudes and perceptions of the program structure. The first part of the interview consisted of having each nurse rate on a five-point scale 16 major aspects of the mental hygiene program. The interviewer presented verbally each aspect to be rated, and when requested by the nurses would clarify or amplify the meaning of the specific

---

*Received in the Editorial Office on March 15, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

program variable. The points of the program which were rated were developed with the assistance of one of the mental hygiene consultants and represented the major aspects of the entire program and its structure. The rating scale ranged graphically from excellent, 1, to poor or unsatisfactory, 5, with gradations in between reflecting degrees of satisfactoriness and need for improvement. Split ratings were not acceptable; each subject had to make a definite evaluation of 1, 2, 3, 4, or 5.

The remaining part of the interview was structured to elicit the attitudes, reflections, and evaluations of the nurses in relation to the following components of the mental hygiene program: (*a*) the most desirable content; (*b*) the most effective presentation methods; (*c*) the most functional professional personnel and their rôles; (*d*) positive and negative aspects of the program; (*e*) the specialized use of the mental health consultants; (*f*) the perceived orientation of the program; (*g*) public health nursing goals in mental health education; (*h*) desirability and need for the program; and (*i*) the value of increased clinical service as part of mental health consultation.

The interview was designed in a structured manner in order to permit systematic interviewing techniques and standardized recording in an objective manner. An interviewing recording form was developed and all interviewers were trained in the handling of this form, so as to permit a reasonably high degree of standardization and consistency. Although the interview was content-structured, it was to a high degree non-directive in regard to nurse response and did not involve any "forced choice" responses; thus the material obtained was in the greatest part dependent upon the productivity of the nurse subject and particularly upon her ability to deal with both conceptual and segmental aspects of mental health in public health nursing and in-service mental health education.

In this respect the nurses' responsivity represents an indirect reflection upon her grasp of the nuclear meaning of mental health and its rôle in public health nursing.

Within the framework of the five-point scale employed in the current evaluation, this nursing group tends to see the content of the program, the quality of teaching ability, and the techniques of consultation as being slightly less than satisfactory. The standard deviations reflect the general consistency of this evaluation, though the coefficients of variability show significant differences in the patterning of the evaluations of certain groups of nurses on these items.

Individual case conferences were generally regarded as being very valuable and were considered more than satisfactory in calibre. Variability for this item was reasonably high, 13.2, and reflects the tendency for this service to be decidedly more valuable for some nurses than for others.

Group conferences were thought to be much less useful than the individual conferences and the entire group with almost no variability, $V = 2.8$, experienced considerable dissatisfaction with this technique and indicated it needed a good deal of improvement.

In appraising the conceptual orientation of the program, the nurses felt that little information had been presented regarding the various schools of thought within the framework of mental hygiene. Their evaluations reflected strong dissatisfaction with this omission and a high degree of agreement in this regard among all the nurses.

Among the aspects of the program that the public health nurses thought needed improvement, no aspect received greater criticism than the presentation of material relating to the nature, meaning, and use of psychotherapeutic procedures.

The vast majority of the nurses felt the program presented mental hygiene concepts in a very satisfactory and objective manner which lent itself to functional incorporation into nursing rôles. Similarly, it was felt that the program effectively defined the rôle of the public health nurse in mental health areas.

A less than satisfactory rating was given to the estimate of how well this program had reflected the needs and requests of the nurses themselves. The dispersion of ratings depicts considerable disagreement as to the extent to which the program failed to measure up to this criterion, though the trend is clearly toward a negative assessment.

The nurses thought that the greatest need for improvement was to be found in the following areas: topical content of the program, the teaching ability and communication techniques of the consultants; group conferences; the incorporation of public health nurse needs and goals within the program, as well as the development of an objective measure of individual and group achievement.

The nurses were able to outline the various consultation and/or teaching services they thought essential to a mental hygiene program and were able to indicate the rôles they felt each discipline might assume most productively, assuming all were available.

With respect to general case consultation, both group and individual, 95 per cent of the nurses thought the public health nurse consultant would be most helpful in assuming this part of the program. The use of the social worker, psychologist, and psychiatrist in this capacity was seen as being much more limited. The public health nurse consultant was seen as generally integrating the total program and assuming responsibility for its general interpretation. The psychologist and psychiatrist were seen as playing a major rôle in presenting psychopathological and psychodiagnostic material. This was also true for the presentation of psychodynamics and developmental psychology. The rôle of the psychiatric social worker in handling problems of case referral and community resources was seen as being the most significant contribution by this discipline in the program. Other functions of the social worker included family visits and information of family relationships as well as information on clinic procedures and interviewing techniques. Additional services from the psychiatrist included the presentation of principles of psychotherapy and consultation on specific problems (along with the nurse and social work consultants). The psychologist was also viewed as being helpful by having direct consultations with patients and presenting aspects of current research.

Although the major purpose of this study has been to obtain an objective evaluation of an in-service mental health program by the participating public health nurses, concomitant and allied factors also present themselves. For one, the spontaneous productivity of the subjects of this study was clinically as well as statistically impressive. Dealing with a highly variable topic and one with which there was a good deal of personal involvement, the nurses were able to delimit the salient variables in this program and to evaluate them constructively.

The public health nurses in this group appear to be aware of their professional rôles as nurses and somewhat confused as to how well the mental health program was able to structure its function within this framework. The expressed dissatisfaction with the topical content of the program and the definite suggestions for future inclusion and expansion reflect interest in mental health generally, and a wish to play a more intensive rôle professionally in this area.

Within the various classification groups of this nurse population there were noted significant variations in both the evaluation of the group program and attitudes toward it. The youngest and the oldest groups, as well as the least experienced and the most experienced appeared to be critical of the program.

These groups are also the most homogeneous of these classification groups. This trend is clear throughout the study: as a group of nurses tend to be less variable in their attitude and evaluation, they tend to be more critical and less satisfied. This may imply that relative success and failure in goal attainment within the program studied here has been significantly influenced by two things: (*a*) the failure to take into account major differences among the nurses themselves both with respect to their psychological and occupational orientation, and (*b*) the isolation and understanding of the meaning and direction of nurse variation and the development of consultation methods designed to best meet their differences.

CONCLUSIONS

1. The nurses involved in this mental health program appear to have achieved an objective concept of the value of mental health within public health nursing.
2. The nurses express more critical attitudes toward the total program relationship than toward any of the component factors within the program.
3. Nurses' evaluations of mental health program in which they have been participants reveal objective satisfactions and dissatisfactions which lend themselves to constructive suggestions for improvement.
4. The most satisfactory parts of the mental health program from the nurses' point of view: (*a*) Individual case conferences; (*b*) opportunity for individual participation; (*c*) general flexibility of the consultation relationship; and (*d*) stimulation of interest in mental health.
5. The strongest dissatisfactions were expressed in relation to: (*a*) topical content; (*b*) the teaching abilities and communication techniques of the consultants; (*c*) group conferences; (*d*) the lack of well defined program goals consistent with public health nursing needs; and (*e*) the lack of objective measures of achievement within the program.
6. The nurses see a need for the inclusion of psychiatric and psychological personnel as well as public health nursing and social work consultants.
7. Specialized rôles are seen for each of the above personnel when all are available.
8. There is an indicated need for a change in the topical content to be more consistent with the expressed needs of the nurses.
9. Significant differences in reaction to the mental health program is evidenced among certain classifications of nurses and mental health consultation techniques appear to require considerable variability in order to work effectively with what appear to be significant psychological differences within a public health nursing population.
10. In general, as a group of nurses tend to be less variable in their reaction to the mental health program their evaluations tend to be more criti-

cal and lower. As variability increases the evaluations tend to be more favorable.

11. The need to further study the possible psychological and occupational nature of public health nursing variability would appear to be important in extending the scope of mental health consultation and application.

SUMMARY

This paper reports a study of nurses' reactions to an in-service educational program in mental health of two years' duration. The results are presented in the hope that they might facilitate the structuring of such programs to more adequately meet the expressed needs of nurses, until such time as we may be able to contribute more knowledge as to the methods and techniques that might facilitate learning based on the psychological data as to the actual learning characteristics of the nursing group. As such, this is the first step in research into the nature of learning characteristics of public health nurses and their mental health functions.

REFERENCE

1. KAZAN, A. T., OSTROW, E., CUMINGS, R., & KLINE, M. V. Teaching mental hygiene: A problem in resistance. *Psychiat. Quart.*, (in press).

Westchester County Department of Health
County Office Building
White Plains, New York

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 201-206.

SOME FACTORS IN THE PRODUCTION OF GAMMA MOVEMENT*

Department of Psychology, Michigan State College

S. HOWARD BARTLEY AND FRANK R. WILKINSON

A. PROBLEM

The apparent movement observed when all or a portion of the visual field is suddenly raised or lowered in intensity is a highly manipulable phenomenon. In order to come to grips with it in the laboratory, one must choose a limited number of its aspects to deal with at a time. The first step is deciding upon whether to use a restricted target or manipulate the intensity of the whole field as a unit. A common choice has consisted in using targets subtending relatively small visual angles. This is not generally a simplification. Targets of restricted area scatter effective amounts of stray light on all the retina outside the target image (1, 4). Although this stray light often participates in producing the experimental outcomes, its existence is generally overlooked and its rôle not taken into account in explanatory interpretations of results. It has been shown that certain explanations of this sort have been erroneous on this account (2).

One idea that has been put forth to account for apparent movement of a target border when the target emerges on a dark background makes movement a feature of the emergence of a figure on a ground (2). The implication, if not the assertion, here, is that movement is an essential and integral feature of the psychological figure-ground phenomenon, and its conditions are not to be looked for at the peripheral end of the optic pathway. This assertion was challenged and refuted by Bartley in 1941 (3), when he reported that whether or not apparent movement of a target border occurred depended upon the visual angle subtended by the target. Targets subtending small visual angles manifested border movement, whereas targets of greater visual angle did not, but, instead, exhibited movements in regions within their borders. No extended study of the conditions for the production of this movement (gamma movement) was made at that time. It did appear, however, that the emergence of movement had some dependence upon the retinal region involved. Ease of production of movement seemed to

---

*Received in the Editorial Office on April 25, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

be greater near the center of the retina than further toward the periphery, at least under the more common stimulus conditions.

Recently the subject again became of interest, and therefore an examination of some of the possible conditions for border movement was undertaken. In the study, two variables were involved: target subtense (or its corresponding area), and target intensity.

B. METHOD AND PROCEDURE

A series of six disks, 1, 1/16, 1½, 2⅛, 3, 4½, and 6 inches in diameter were used as the targets. These were opal glass-covered openings in a screen produced by a series of interchangeable slides, each with its appropriate disk opening. The targets were viewed from each of two distances: 17.0 inches, and 66.8 inches. At the greater distance, the targets subtended visual angles of 1°, 1°26′, 2°, 2°34′, 3°38′, and 5°8′. At the shorter distance they subtended angles of 3°36′, 5°04′, 7°12′, 10°6′, 14°12′, and 20°. The two ranges overlapped, but in essence formed a series in which the *area* of each succeeding target was twice that of the former. Hence the targets formed a series of relative areas of 1, 2, 4, 8, 16, 32, 64, 128, 256, and 512, beginning with a subtense of 1 degree.

A revolving disk with an open sector produced the flashes that, in one case, had an over-all duration of 48 msec.; 16 msec. for the rise to maximum; 16 msec. duration at maximum, and 16 msec. for decline. In the second part of the experiment, the pulse duration was increased to an over-all of 144 msec. with the same time for rise and decline as the already-mentioned short pulse. In the third part of the experiment, the pulse was tapered so that 128 msec. were used in the rise and 16 msec. in the decline immediately following the rise. The pulses were delivered at a rate of one every 1200 msec. The authors were the observers in the experiment. The procedure consisted in going from the intensities which did not produce border movement to those at which border movement just became indubitably apparent. At some intensities, the general appearance of the target, when the borders were not specifically attended to, was that of expanding when the target emerged. But when the attention was directed to the borders, the movement was seen actually not to involve the border. We make this caution, for it would be most probable for anyone wishing to repeat our experiment to take the over-all expansion appearance of the target to indicate that the borders moved.

C. RESULTS AND CONCLUSIONS

The data for one of the two observers are plotted in Figure 1. Curve *A* is for the short light pulse; *B* for the long untapered pulse; and *C* for the long tapered pulse. Curves *A* and *C* were not too difficult to obtain, but the conditions involved in *B* were such as to produce the two-flash phenomenon (5). When this phenomenon was present, it was virtually impossible to obtain readings which meant anything for the purpose at hand. The

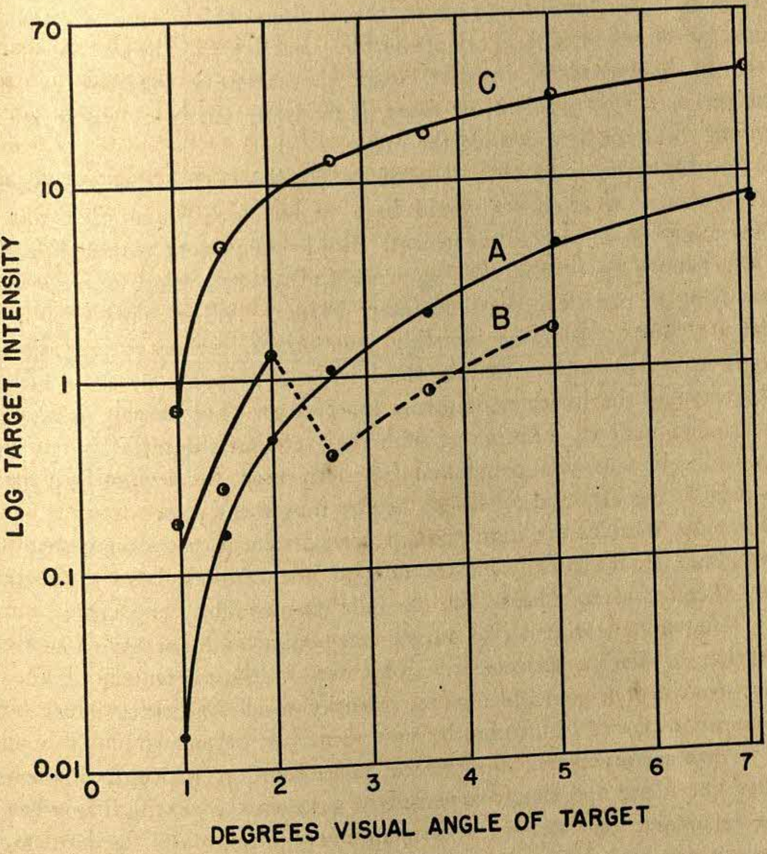


FIGURE 1

DIAGRAM SHOWING THE RELATION BETWEEN TARGET SIZE AND INTENSITY FOR THRESHOLD OF BORDER MOVEMENT AS THE TARGET APPEARS

Curve *A* is for a target 48-msec. long, with 16-msec. onset and decay. Curve *B* is for a target 144-msec. long, with same onsets and decay as short target. Curve *C* is for a target whose onset is 128 msec., and immediate decay is 16 msec. Conditions in Curve *B* give rise to the two-flash phenomenon, and thus disrupt the comparisons with *A* and *C*.

alternative chosen was to drop below the intensity producing the double flash and take border readings there. This accounts for the break and sudden drop in the level of the curve. The least difficult readings were those obtained for Curve *A* with the short pulse of light. Even they were somewhat subtle when large visual angles were involved, this accounting for the fact that our curves do not extend to 20 degrees, the limit initially intended.

It will be obvious from the curves that as visual angle is increased, the intensity level for threshold border motion rises. This calls for an explanation. In accounting for border movement, the findings which have accounted for gamma movement in general are applicable. Bartley (2) showed that gamma movement is dependent, in part, upon the intensity gradient of retinal illumination. This gradient, or taper, is made up partly of the image, and partly of the stray light outside the image.

It would be expected that to produce conditions for the perception of border movement, gradients would have to be more pronounced than for the mere impression of field movement. Border movement would depend upon an effective spatio-temporal sequence in the activation of retinal elements lying at and near the final image border. When some parts of the retina are illuminated more intensely than others, retinal activity in those parts is instituted earlier; that is, the latency of response of those elements is shorter than the latencies of others involved in progressively weaker and weaker illumination. This, along with the fact that the retina is not homogeneous in its sense-cell population has been used to account for gamma movement in general and applies to border movement also.

Our more specific task is to account for why large targets must be more intense than small centrally located ones to just manifest border movement. From what has already been said, the following could be expected:

1. That very low levels of target intensity would not give rise to the perception of border movement. This we found to be the case. No border movement is seen until target intensity reaches a critical level. This is true for all sizes of targets.

2. That as target area is increased, target borders will fall further and further out along the center-to-periphery gradient (taper). Therefore, in order to involve the same or equivalent over-all sense cell activity, higher intensities are needed than for small central targets. In this connection, it might be postulated that the crucial sense cells for the function with which we are here dealing are the cones. In support of this, we have the work on gamma movement just mentioned (2, 3), and in particular the reversal of gamma movement by using very large targets whose intensities tapered upward from center to edge. Cone populations density decreases as distance

from fovea increases, whereas the opposite is true for rods up to a certain point (18 degrees in the temporal half of the retina). Possibly rod system function is slower than that of the cone system (7), hence the density of concentration of cones forms the crucial determiner of time characteristics of retinal response under such conditions as ours. Increasing intensity of retinal illumination in the larger targets would tend to raise the over-all

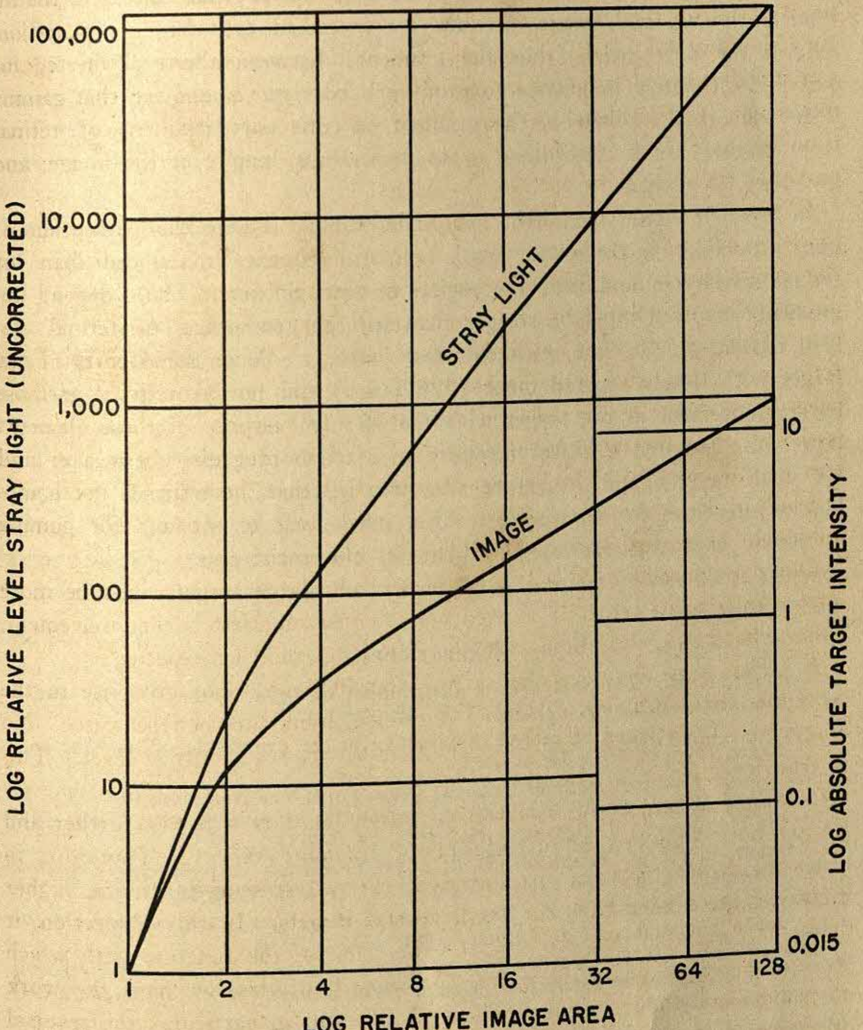


FIGURE 2
A DIAGRAM SHOWING THAT INTENSITY OF STRAY LIGHT INCREASES FASTER THAN THE INTENSITY OF THE TARGET IMAGE: THIS IS FOR CONDITIONS OF CURVE *A*, IN FIGURE 1

activity to a level at which gradients could function in producing border movement. Raising intensity beyond this point would tend to bring intensity differences between image and stray light to a minimum or even to a vanishing point. Observations are in line with this. Once target intensities become very high, the sharpness of borders is lost, both as the target emerges and as it is observed in a static state.

3. That intensity of stray light increases faster than target intensity when both area and target intensity are increased (see Figure 2). This reduces the differential (thus the gradient) between the two levels and makes it ultimately impossible to produce border movement for this reason, if for no other. Likewise, the gradient of cone population vanishes with large targets for it diminishes to ineffectiveness long before the extreme periphery is reached.

In brief, it would seem that border movement is dependent upon conditions providing for the proper spatio-temporal sequence in the institution of retinal activity in and near the region of final border. The slope of the gradients involved must be steeper than for the production of mere diffuse field movement. In line with this last point, we have the fact that our target with slowly tapered onset (128 msec.) did not provoke as striking border movement as the target with the 16-msec. onset. The long-tapered target also required a higher intensity to reach border movement threshold for small targets, and, as target area was increased, continued to require higher intensities for thresholds. Also, border movement was lost earlier as target area was increased. That is, the satisfactory spatio-temporal gradient could not be maintained as far out on the retina as with targets of more abrupt onset.

REFERENCES

1. BARTLEY, S. H. The comparative distribution of light in the stimulus and on the retina. *J. Comp. Psychol.,* 1935, **19**, 149-154.
2. ———. The relation of retinal illumination to the experience of movement. *J. Exper. Psychol.,* 1936, **19**, 475-485.
3. ———. Vision: A study of its basis. New York: Van Nostrand, 1941. Pp. 350.
4. ———, & FRY, G. A. An indirect method for measuring stray light within the human eye. *J. Opt. Soc. Am.,* 1935, **24**, 342-347.
5. BARTLEY, S. H., & WILKINSON, F. R. Certain factors in producing complexity of response to a single pulse of light. *J. of Psychol.,* 1953, **35**, 299-306.
6. FRY, G. A., & BARTLEY, S. H. The relation of stray light in the eye to the retinal action potential. *A. J. Physiol.,* 1935, **111**, 335-340.
7. POFFENBERGER, A. T. Reaction time to retinal stimulation, with special reference to the time lost in conduction through nerve centers. *Arch. of Psychol.,* 1912, No. 23. Pp. 73.

Department of Psychology
Michigan State College
East Lansing, Michigan

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 207-241.

MODERN THEORIES OF OLFACTION: A CRITICAL REVIEW*

Department of Psychology, University of California

F. NOWELL JONES AND MARGARET HUBBARD JONES**

A. INTRODUCTION

Following the flurry of activity sparked by Zwaardemaker and Henning during the first decades of this century, interest in olfaction sank to a relatively low level. It is true that there is a sizable subsequent bibliography, but it is not nearly as extensive as those in vision and audition. Very recently, interest in the "chemical" senses has again quickened, and the time has come to appraise the current status of theory in the field of olfaction. That there is not appreciation by psychologists in general of what recent work there is, is evident in the current textbooks, handbooks, and histories of psychology. Apparently knowledge in this field has not progressed since Henning, if the textbooks are to be believed, and his contribution to olfactory theory, to knowledge of basic qualities and classification of odors are treated as good modern doctrine and scientifically sound. Woodworth's *Experimental Psychology* devotes some space to olfaction (19 pages), but the latest reference is 1930, and there are only three subsequent to 1925, whereas in vision and audition the latest references are 1938, with many subsequent to 1930 (137). Again, Boring's *History of Sensation and Perception,* although rather thorough in other areas, leaves olfaction with Henning, and gives no clues to the current problems and no real insights into the difficulties which beset experimentation and theorizing in this field (23). Of the three popular Physiological Psychologies, only Morgan and Stellar attempt an up-to-date summary of the status of olfaction and theirs is rather brief (94). A fairly recent attempt at a thorough review of olfaction is contained in Moncrieff's *The Chemical Senses,* but here the approach is not psychological and much of the psychophysical work has been omitted (91). Further, some important work, both experimental and theoretical, has appeared since the publication of the book. Finally, insufficient space was allotted the chemical senses in the *Handbook of Experimental Psychology,* and the treatment there is perforce somewhat summary (113). Haagen-

---

*Received in the Editorial Office on April 27, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

**This review was prepared as part of the work of Project NR140-008, Contract N6onr-27515, between the University of California and the Office of Naval Research.

Smit has written the most modern summary (59), but it is written for the layman and is not sufficiently detailed for the professional psychologist. There seems to be no source to which the general psychologist can go for a thorough briefing on modern research.[1] This article is an attempt to provide a critical review of modern theories, based upon current knowledge, together with an evaluation of current possibilities for fruitful hypotheses.

Although we interpret "modern" to mean within the last 10 or 15 years, we shall include some authors who immediately followed Henning, both because of the apparent lack of awareness of their contributions and because it will provide some continuity for an on-going tale. Every effort has been made to make the review complete, but, since no bibliography is ever complete, we have undoubtedly missed some pertinent references, and there are a few references which were not obtainable.

Let us first be explicit about Henning's theory, without reviewing it in detail. (For a detailed review, cf. Boring (23), pp. 445-449, Woodworth (137), pp. 482-495, Henning (67), p. 205, Moncrieff (91), pp. 168-170, 173-235, 302-303.) A gross chemical theory, such as Henning's, in which particular groups of atoms are responsible for particular odors, is now considered untenable. Cohn (34) pointed this out in 1924 [see also Tschirch (131, 132)], and even Henning himself later largely lost faith in odorophores. The reasons for abandoning this type of theory are many and, in the aggregate, compelling. Among them are the following: (*a*) some substances change odor quality with concentration; (*b*) very similar odors may be caused by compounds which differ greatly in structure; (*c*) stereoisomers may have different odors; (*d*) compounds which are very similar chemically may have quite different odors; (*e*) it has not been possible to predict the odor of a compound, or even whether it will be odorous, from the chemical structure.

Before proceeding to a discussion of actual theories, it might be well to discuss some of the conditions which have been deemed necessary, though not sufficient, for odor. Cohn studied odorous substances, and pointed out what he considered to be some of the requisite properties of odorous substances which are still widely quoted: (*a*) a substance must be, to some significant degree, volatile at room temperature; (*b*) it must be at least a trace soluble in water; (*c*) it must be within certain limits as to molecular weight (not too high); (*d*) it must be to some extent soluble in lipoids.

---

[1] There has recently (1952) appeared a reasonably extensive bibliography of olfaction (7). However, the latest reference is 1947, it is unannotated, and the classifications are often misleading.

Adsorbability on the mucous membrane undoubtedly plays some rôle, he thought (34). These dicta are not as useful as they seem, in spite of their great popularity. Volatility, it is true, is necessary, because molecules of the substance must get into the nose, but it is not possible to set a definite limit, and everything is to some extent volatile. Pure gum rubber, for example, is not considered to be particularly volatile, but it is nonetheless rather odorous—an unfortunate fact which plagues builders of olfactometers. Solubility in water is another teaser, for everything is a trace soluble in water, and no definite limit has ever been set. This requirement is based on the assumption that the olfactory membrane is covered with mucus and that all stimuli must first be dissolved in it before they can reach the sensory cells. Although this appears possible, it is not quite clear whether the olfactory hairs may not protrude slightly from the surface of the mucus (cf. 73). To support the negative argument, there are cases where a trace of water solubility must be accepted as sufficient. Lipoid solubility has long been on a more solid footing, with few criticisms. But this requirement seems to have been dealt a serious blow by Masaji (86), who found that when trypan blue, which is lipoid insoluble, is used on the surface of the olfactory epithelium, it *selectively* colors the olfactory cells—in 10 minutes the nucleus is deeply tinged with color. It is apparently possible for a substance to enter the sensory cells rapidly even without lipoid solubility. In addition, it should be pointed out that stimulation may be a surface phenomenon, and not involve penetration of the receptor at all. That molecular weight must be within certain limits has emerged from empirical work, but is not a precise requirement. Adsorbability on the mucous membrane is a suggestion which cannot be disposed of on the basis of present evidence, but it is contradictory to the dictum of lipoid solubility and thus introduces inconsistency into Cohn's discussion. More recently Morimura (108) has stated that the mechanism of smell is dependent upon the vapor pressure of odorous substances, thus reaffirming Cohn's requirement of volatility. But again we find that vapor pressure, or volatility, though important, is not a sufficient condition, and knowledge of vapor pressure alone does not enable one to predict either the quality of odor or, indeed, whether the substance will be odorous.

These general conditions have not, thus far, proved to be very enlightening, and one turns, therefore, to the theories which have been proposed to account for the meager facts in the field for some enlightenment on the nature of olfaction. There are two major problems to be faced by a complete theory of olfaction: one is the problem of the nature of the stimulus—how the substance actually accomplishes the change in the sensory cell which initiates the nerve impulse; what is the nature of the influence of the odorous molecule upon the receptor cell? The other concerns the problem of olfactory qualities—what accounts for the qualitative differences among odors; how may the vast array of qualitatively different odors be explained? In the discussion of theories which follows, it will be discovered that most theories are designed to answer one or the other of these problems, seldom both. An occasional theory may be stretched to cover both problems, but it is usually more satisfactory in one respect than in the other. The first of these two problems is basic. A really accurate answer to the second, however, would be of considerable value in suggesting answers to the first.

We have attempted to divide the theories into groups, for convenience in discussion. We have no wish to defend the particular classification, which is purely heuristic, nor to claim that all theories placed in a particular group are of equal merit or even necessarily subject to exactly the same criticisms. But there appears to be sufficient similarity among theories of a group to simplify somewhat the discussion. We have parceled the theories out into seven groups: (*a*) radiation and vibration theories, in which either the actual stimulus or the qualitative differences are thought to be in some way dependent upon molecular vibrations of one sort or another; (*b*) mechanical theories, in which the stimulus is conceived of as a mechanical force; (*c*) stimulus pattern theories, in which qualities are determined by patterns of stimulation; (*d*) "chemical" theories (for want of a more appropriate name), in which some sort of effect, best described as chemical in nature, is considered the stimulus; (*e*) steric theories, in which the characteristics of odorous substances are ascribed to the shape and dimensions of the molecule; (*f*) phase-boundary theories, in which an effect occurring at some sort of phase boundary is the stimulus; and (*g*) enzyme theories, in which odorous substances are thought to influence enzyme systems within the receptors. The last categories are, of course, all "chemical," but they imply different emphases upon the different aspects of receptor function.

B. RADIATION AND VIBRATION THEORIES

Among theories of this type we find both general kinds of theories, those which deal with the stimulus and those which deal with qualities. One theory of great historical interest is Teudt's, both because it was proposed so early, and because it is one of the few theories requiring action-at-a-distance (128, 129). Teudt felt that some phenomena could not be accounted for by direct-contact theories and therefore tried his hand at another sort. His theory states that electron vibrations in the atom or molecule set up vibrations in the surrounding medium, these in turn being reinforced by resonance. The olfactory nerves are assumed to have similar electrical vibrations, different in different nerves, and molecules of odorous substance will reinforce those whose period of vibration corresponds to their own. Different rates of vibration are perceived as different qualities of odor. Both the mechanism of stimulation and of qualitative differentiation are taken care of by this theory, which, although it was conceived before any very precise knowledge of intramolecular energies was available, and thus of necessity deals in generalities, is not unlike some modern theories in that the resonance of the molecule is thought to be of basic importance, an hypothesis for which Teudt was ridiculed by Heller (63, 64, 65, 66), who himself held a theory of direct chemical reaction.

Heyninx (70) developed an old discarded theory of Zwaardemaker's that molecular vibrations in the ultraviolet range exhibited by odorous substances are responsible for odors. He believed that only some such physical theory would fit the facts, and further argued that the diameter of grains of pigment in the olfactory membrane was such ($.30\mu$ and less) that the vibrations must lie in the ultraviolet, rather than in any other region. His investigations convinced him that all odorous substances have absorption bands between .20 and $.36\mu$ and that a relationship exists between wavelength and quality of odor such that substances which absorb light of wavelength of .20 to $.22\mu$ smell ethereal, those which absorb at .26 to $.28\mu$ smell burned, etc. Such theories as this one are better equipped to handle the problem of qualities than that of the stimulus. Although the theory seems to have an explanation of the stimulus, vibration of certain frequencies, there is no clue as to how this vibration affects the sense cell, so that the real problem of the stimulus is not answered. Furthermore, scrutiny of the basic tenets of the theory reveals that the relationship between ultraviolet absorption and odor is not a perfect one. Niccolini (100), for example, found that of five synthetic coumarins which absorbed ultraviolet light uniformly, some were odorous and some were not. Riccitelli (117) has suggested a similar theory, placing emphasis on the size (diameter?) of the grains of pigment found in the olfactory membrane. Stefanini (126) has espoused Heyninx's theory, but seeks to explain odor by the assumption of three primaries: $.310\mu$, putrid; $.250\mu$, spicy; $.210\mu$, ethereal. Although many people have assumed that depth of pigmentation is correlated with olfactory acuity, Le Magnen and Rapaport (82) were unable to find any evidence that this is so, or even that the pigment is associated with the receptor cells, a conclusion generally agreed upon in histology (cf. 87). Ultraviolet theories are thus not tenable today.

One of the more famous theories in this category is that of Dyson (46, 47). He believed that certain internal vibrations of the molecule, called osmic frequencies, are responsible for odor, and that these intramolecular vibrations may be measured by utilizing the Raman spectrum (cf. Hibben, 71). Every substance has a Raman spectrum which is more or less unique, representing the intramolecular energy relationships peculiar to that substance. Raman spectra are complex—a series of lines rather than a single line—and all odors are thus conceived to be complex. No simple odor, analogous to homogeneous light, is possible. The osmic frequency range is assumed by Dyson to be correlated with Raman shifts occurring between 1,400 and 3,500 $\Delta\bar{\nu}$. The question arises as to whether such a scheme can account for different nuances among odors of the same general type. Dyson claims that geraniol and its stereoisomer, nerol, show slightly different frequencies, and that the intensities of the Raman lines differ for d- and l-pinene. The following examples will suffice to indicate the magnitude of the differences which Dyson believes to be sufficient to account for the differences in odor (Table 1).

TABLE 1

	Prominent Raman lines			
Cis-2-pentene	1,248	1,266	1,375	1,658
Trans-2-pentene	1,298	1,313	1,378	1,674

Dyson claims, in support of his theory, that many olfactory phenomena can be accounted for: differences in odor of stereoisomers, change of odor quality with concentration, similarity in odor of compounds which are chemically different. Dyson also attempts to explain compensation by the interference of two neighboring frequencies with one another. Although the latter argument is not very compelling, it is not clear whether a theory should be bound by this requirement. There are no really adequate data on compensation. It is only the older writers who claim to have made observations on the topic, and they neither agree, nor provide data, nor report adequate control of stimulus or observational conditions.

This theory makes no attempt to explain precisely how these intramolecular vibrations affect the olfactory cells; it is a theory which deals with the basis of qualitative differences, not the mechanism of stimulation. Other than the fact that there is no bridge between the intramolecular vibrations and the receptors, and no hint of how different frequencies might be converted into qualitative differences by means of nerve impulses, there are two specific criticisms of theories of this type. Saksena (120) has examined the Raman spectrum of very carefully prepared glycerin and finds that it contains several prominent lines within the "osmic frequencies," namely at 1,466, 2,880, 2,955, and 3,202. Since it is sufficiently volatile and lipoid-soluble, Saksena says, it should be odorous, but it isn't. Actually, it is not very lipoid soluble, and Saksena's point may not be too strong (but see our discussion of Cohn's requirements).

In criticism of this theory, Young, Pletcher, and Wright (138) have argued that: (*a*) there is no evidence that the odor of any isotopically substituted molecule differs from that of its ordinary counterpart, but the infra-red (and therefore also the Raman) spectra differ, and (*b*) the d- and l-forms of a few optical isomers (particularly those cited by v. Braun and his associates (25, 26)) have different odors but identical infra-red (or Raman) spectra (barring the effect of polarized light). Some recent work by Doll and Bournot (41) indicates that this problem is not a simple one, however. They found it extremely difficult to prepare a compound from different sources and have the samples smell the same even when the physical constants were identical. They finally succeeded in preparing l-menthol from the racemic form and from oil of pennyroyal, and d-menthol from the racemic form and from geranium oil; their experienced observers could not tell the two l-samples apart, nor the two d-samples. Then threshold measurements were made and the d- and l-forms were observed to have identical thresholds and to be indistinguishable in quality. This would *seem* to indicate that isomers having identical infra-red or Raman spectra do smell alike—or at least that d- and l-menthol do, and others might also, if the compounds are prepared with sufficient care. However, Doll and Bournot found that at high concentrations l-menthol is much more intense in odor than d-menthol, and it also has a lower threshold for the cold receptors, so apparently the physiological effects are not identical even when the physical constants are identical. Nevertheless, it is such a general rule that the d- and l-forms of substances differ in biological activity that the work of Doll and Bournot should not be overgeneralized.

One of the difficulties which beset theorizing becomes apparent in discussing this theory, as well as others to follow: the demands for precise physico-chemical information about a large number of odorous substances far exceed the available data. It is not possible to choose a number of substances of certain odor characteristics and then merely look up all the desired physico-chemical data A second stumbling block to accurate information about the relationships between odor and chemical structure is the fact, now fairly well recognized, that ordinary standards of chemical purity are not sufficient for odor control (cf. Doll and Bournot, above); the sense of smell is more sensitive than are chemical tests, and this introduces a factor very difficult of control.

Infra-red radiation has been mentioned as the factor in odor many times. Such theories are among the oldest type: Faraday and Tyndall both pointed out a relationship between infra-red absorption bands and odor. Henry (cf. 60 for an available secondary source) presented a true olfactory theory based upon infra-red radiation, in which odorous substances caused changes in radiation from the olfactory cleft. Beck and Miles recently have proposed a theory very similar to Henry's (20, 89, 90), and Dyson's earlier theory is likewise similar (45). Their theory has been concerned with possible receptor mechanisms rather than with the correlation of particular odors with specific infra-red absorption bands. They have sought to support their theory with behavioral studies of bees and roaches, but the results appear equivocal.[2] The theory states that the receptor elements radiate selectively, dependent upon size and shape, and when any odorous substance comes within the radiation field of the receptors, it causes them to lose energy because of the infra-red absorption characteristics of the substance.[3] This loss in energy initiates the nerve impulse. Because the receptors radiate selectively and the substances presumably have different infra-red absorption spectra, certain receptors will be stimulated by some substances and certain others by other substances. There is no unequivocal evidence in favor of the theory, and the criticisms of Young, Pletcher, and Wright (138, vide supra) seem to indicate that both the Raman and infra-red spectrum theories are inadequate. Recently Beck (19) has made some observations on optical isomers (d- and l-citronellol, d- and l-limonene, and d- and l-carvone) and, in line with some previous work (25, 26, 124), has found that they may smell different qualitatively although their infra-red absorption spectra are identical (but cf. Doll and Bournot, 41). Beck attempts to rescue his theory by postulating that the odorous substance is adsorbed as a monomolecular film on the olfactory cells and that the cells "contain an optically active compound which reacts with some entantiomorphs forming a loosely bound racemic compound. The receptors then cool off by radiating through the capillary film" (19, p. 661). This modification will be difficult of proof also, particularly in view of the fact that it must be shown that the olfactory receptors are required for infra-red discrimination, that the responses obtained in the insect work are not due to visual or temperature discrimination. Further, Beck has now conceded that some substances may stimulate by direct chemical reaction. This would greatly complicate the attempt to correlate stimulus variables and odor.[4]

---

[2] In Beck and Miles' study of bees, the N is small and the results are not treated statistically. It is quite possible that the differences found are not statistically significant. But an even greater difficulty arises from lack of experimental controls. Even if bees are shown to go more frequently to the infra-red transmitting window, the reason for their doing so would not be obvious. Since controls were not instituted, either with no odorant (honey) behind an infra-red window or with the olfactory receptors of bees removed, it is not clear that the bees would not be responding to infra-red radiation *per se*, rather than through the medium of olfaction.

[3] Beck assumes that the diameters of the olfactory receptors determine the wavelengths at which the receptors radiate. This may or may not be true, and is apparently based on analogy with macroscopic antennae, which may not be applicable to such minute wavelengths, where molecular structure becomes important.

Pressler (116) has seized this solution, and suggests that there may be two olfactory mechanisms, one chemical "as in taste," the other vibrational as in Beck and Miles' theory. Nothing new is offered here, except the thought that olfactory problems can be solved by tossing them into the pot with taste.[5]

Müller (96) has proposed a dipole theory in which the differences in odors are due to the differences in the dipole moments of odorous molecules. The odorous molecule is said to "set up an irritation" at the moment of contact, then to become neutralized and carried away from the receptor surface. This theory, its author contends, would account for the often-mentioned discontinuity of odor in expiration, for adaptation due to the piling up of neutralized molecules which are not carried away by the mucus fast enough, and for changes in odor quality with concentration, the last by assuming association of molecules in highly concentrated solutions which would alter the effective moments. Unfortunately, there are odorous compounds which have no dipole moment (benzene is a good example), and although Müller recognizes this, he deals with it ineffectively, calling such substances "pseudo-aromatics" and the influence they exert upon the olfactory system an "abnormal influence." This is a theory of the way in which the receptors are stimulated; presumably qualitative differences might be ascribed to different values of the dipole moment. Müller does not argue very effectively for this kind of theory. It is possible that dipole moment may be one of the important aspects of the molecule, rather than the only one. Like most theories, Müller's is too "molecule-bound," with little attention given to the receptor cell. Müller later investigated the fluorescence of many odorous compounds when exposed to ultraviolet light (97), but he found substances too similar to make this characteristic into a theory.

---

[4] A criticism of Beck and Miles apparently originated by Moncrieff (91) states that carbon disulfide and "paraffin" (kerosene?) do not show infra-red absorption. This statement has been repeated by Pfaffmann (113) and Morgan and Stellar (94). This criticism is, of course, unfounded, since these substances do show infra-red absorption (98).

[5] Several comments are in order in this regard: taste and smell are *not* closely associated, as was formerly believed (cf. Börnstein (22)), the receptors are quite different, and there are almost as many theories of taste as there are of smell, none of which is on a very solid footing.

Von Skramlik (125), whose concern is with threshold energies, has attempted to find an explanation for the "tremendous number" of molecules necessary for stimulation, which he says, from older data, is of the order of 10^8 to 10^{17} molecules. Vapor pressure, he thinks, is not the decisive factor, although it plays a part, at least in homologous series. Kinetic energy of the molecule, when calculated, turns out to be too large, but he suggests that possibly the rotation and/or vibration energy may be the important factor. He concludes that we do not yet know enough to determine the minimal energy for olfactory stimulation.

All vibration theories have been criticized by Gavaudan and his collaborators (54). They studied olfactory thresholds, expressed in terms of thermodynamic activity, of aliphatic alcohols, and state that in olfaction the activity threshold shows oscillation, as one ascends the homologous series from C_1 to C_8, whereas in both narcosis and gustation the activity threshold remains almost constant. This suggests to them that the mechanisms of olfaction are rather different from those of narcosis and taste (many theorists assume them to be the same). They believe that the action of the olfactory molecule is more specific than that of gustatory molecules, but that the solution to the problem of the effective stimulus will ultimately be solved in

---

[6] Dethier and Chadwick (30, 31, 38, 39, 40) have some very interesting conclusions, drawn from their work on contact chemoreception. Whether or not these are relevant to the problem at hand is a point of argument. Dethier and Chadwick are of the opinion that chemoreception is much the same throughout the animal kingdom. If this is so, then it is possible that their work might throw some light on the nature of the stimulus for olfaction, even though it was done on insects, and on contact chemoreception. They have attempted to find the chemical and physical properties responsible for the stimulating efficacy of various compounds, and have rejected the following: boiling point, dipole moment, degree of association, melting point, dielectric constants, number of carbon atoms, molecular weight, molecular areas and volumes, oil-water distribution coefficients, activity coefficients, standard free energies, and vapor pressure. "Thus we must conclude either that we have overlooked the significant factors or that we are not dealing with a single process throughout, since the evidence from the series considered individually is convincing that stimulation is in some way related to the thermodynamic properties of the compounds" (31, p. 450). They think that there may be at least a two-phase system operating, and advance the hypothesis that small molecules gain access to the receptors in part through an aqueous phase, whereas the larger aliphatic molecules penetrate chiefly through, or accumulate in a lipoid phase. (Cf. however, Masaji, 86.)

terms of the general kinds of variables which are now used in explaining pharmaco-dynamic and toxic phenomena, not in terms of vibrations of some sort.[6] There does not appear to be any convincing evidence for vibration theories of any sort.

C. MECHANICAL THEORIES

So far as we have been able to discover, only one purely mechanical theory has been proposed. Banerji (16) thinks that air movement in the nose sets the olfactory hairs into vibration, and this vibration is then modulated by odorous particles according to their own weight and momentum. This theory has one observation in its favor (but cf. 21), that odor occurs only when the stream of air is in motion, but odor quality has not been found to be correlated with molecular weight.

D. STIMULUS PATTERN THEORIES

Adrian has proposed rather a unique type of theory to explain olfactory qualities (although not olfactory stimulation). In brief, this theory proposes that different odors stimulate maximally different regions of the olfactory membrane, and that the basis for this differentiation lies in the physical properties of the odors, and the eddy currents in the nasal passages. Although Adrian earlier considered it likely that there were some specific receptors for different types of substances (1, 2), he now believes that spatial patterning, coupled with some temporal differences in time of arousal and decline of response, is sufficient to account for quality (3). These ideas are based on experiments which show that certain "ethereal" odors, e.g., amyl acetate and ether, cause maximal response toward the front of the olfactory bulb in the rabbit, while "oily" odors, specifically benzene and pentane, show maximal activity toward the back (4, 6). It was also found that the "ethereal" odors led to more rapid arousal and decline of activity than did the "oily."

These ideas lead to several points which require discussion. First, it should be pointed out that Hasama (62) did not find spatial patterning, and did find evidence of adaptation, whereas Adrian did not (5). Perhaps further work is needed to settle these issues. Second, Hesse (69) reports that cocainizing different portions of the olfactory membrane does not change perceived odor quality, although he gives no reference. If Hesse is correct, it would seriously embarrass a theory depending on a *place* determination of odor (cf. the analogy with restricted cochlear lesions and tonal gaps). Third, if the structure of the nose is important in determining quality by participating in the "sorting" of odors, then odors in reverse flows of air should be different in quality from the normal. Doll and Bournot (41) find that odors introduced at the choanae have normal qualities, and the present writers' somewhat casual observations confirm these results. Fourth, there must be a preservation of spatial distribution from receptor surface to olfactory bulb, if Adrian's suggestions are correct. Clark and Warwick (33) found no such anatomical relationship in a careful histological study, but Clark (32), subsequent to the announcement of Adrian's theory, reopened the question in a preliminary note, and reported finding the required projection.

There are thus some difficulties with Adrian's theories, whether or not his experimental results are confirmed by others. A more appealing suggestion is that, just as is the case with vision, the various types of receptor are not equally distributed—there are "odor zones" in the olfactory membrane, just as there are "color zones" in the retina.

E. Chemical Theories

The chemical theories are a diverse lot, covering a wide range in time and in responsible factors. Ruzicka (119), although rejecting Zwaardemaker's and Henning's classifications, espoused a chemical theory of olfaction in which direct chemical reaction between the odorous substance and the receptors occurs. Certain chemical radicals—hydroxyl, aldehyde, ketone, carbonester, nitril, nitro, diazoimido—are osmophoric groups, if the substance is soluble in air. Upon access to the olfactory region, they combine with the osmoceptors in the receptor cells. There are thought to be both primary and secondary osmoceptors. Change of quality with adaptation is explained by saturation of all primary osmoceptors, so that the secondary osmoceptors are called into play. When all loci are occupied, complete adaptation would occur. The exact relation between odor and chemical constitution will not be found to be simple, he thinks, but all substances which smell similar must have something in common—some part of each molecule must have a similar effect upon the sensory cells (this after stating that any such system as Henning's must fail because it is based upon "subjective" factors!). But we also find here a suggestion of a steric theory in the statement that it is the "skeleton" of the odorous material which is most important: "Für die Geruchsart ist also am wichtigsten das Skelett des Riechstoffs, durch die Art der Reaction der osmophoren Gruppen mit verschiedenen Osmoceptoren wird in der Regel lediglich die Geruchsnuance bedingt" (119, p. 130). Although Ruzicka states that his theory is a chemical one, and he undoubtedly conceived of the actual stimulus as a direct intervention, there are other aspects to the theory which make it difficult to classify. The theory is primarily a theory of stimulation, with no very specific hypothesis of the exact nature of the change in the receptor. Insofar as it involves a purely chemical reaction between certain kinds of radicals and receptor cells, the same criticisms apply as were pointed out in connection with Henning's theory. As far as the suggestion of steric factors is concerned, the reader is referred to the steric group of theories for a general evaluation.

Woker (135) was impressed with the degree of saturation as a factor in the odor of volatile substances. She thought that the more double or triple bonds a substance possesses the greater the "internal tension" of the molecule and the more intense the odor. To take care of those substances which are undeniably saturated and yet have an odor, she suggested that some element might not be showing its maximum valence. Durrans (2, 43, 44) developed a very similar theory, in which "residual affinities" are considered the source of odor. The oxygen atom, the benzene ring, and multiple bonding are sources of odor which possess such residual affinities. A great deal of the argument rests upon multivalent atoms. In cases where the argument does not fit, it is claimed that the residual affinities are occupied within the molecule or that some atom is operating at other than usual valence. This theory is one of stimulation, no explanation of qualities being offered. It is just another case of argument rather than evidence. It has not been a rewarding hypothesis in the further study of the olfactory mechanisms.

Watson (134) found that amount of depression of surface tension by a substance was fairly closely correlated with its "intensity" of odor (apparently casually determined). No theory of olfaction is offered, but one is reminded of Taylor's theory of taste (127). Watson found an even closer relationship between charcoal adsorption from aqueous solution and "intensity" of odor. No further development of these suggestions is offered, and they can hardly be considered as hypotheses. Reduction in surface tension is also suggested by Moncrieff (vide infra, 92) as a possible mechanism of stimulation, but it should be recognized that the surface tension of the receptor cell is already very low and it is likely that substances interfering with the adsorbed protein layer on the receptor would raise, not lower, the tension (35).

Niccolini (100, 101, 105, 106, 107, 108) thinks that the lack of success in the search for a dependable relation between chemical constitution and odor may be due to looking for too simple relations, and suggests that the oxidation-reduction function may be the important one. He states that: (*a*) odor is associated with anions, (*b*) odor is associated in many series of compounds with reducing power, and (*c*) the receptor is stimulated whenever the oxidation-reduction potential of its environment swings toward the negative side (107). He thinks it would not be possible to account for the diversity of odor of different substances belonging to the same chemical group, nor for the rapid recovery from fatigue, if a chemical combination between the odorous molecules and the protoplasm of the olfactory cells occurred. The "infinite gamut" of perceptible odors is the series of negative oxidation potentials possible in the nasal mucosa.[7, 8] Niccolini's argument here against chemical combination is not at all convincing since it is well known that chemical reactions occur in drug action, and may be very rapid as well as highly selective.

Niccolini also thinks that catalysts may be involved in olfaction. He finds some evidence that thiocyanic acid may be a catalyzer of oxidative reaction. He found it to be decreased in nasal and lachrymatory secretions when eucalyptol was applied, but unchanged in saliva and urine (102, 103, 104).

Recently attempts have been made to verify these hypotheses. Manci (85) determined olfactory thresholds for ketone musk and guaiacol before and after drenching the nasal mucosa with a fog of each of four substances presumed to change the ionic equilibrium of the mucosa. Results of different sign were obtained with those which should augment the concentration of thiocyanic ions and those which should augment the concentration of hydronium ions. Unfortunately the Elsberg technique was used for threshold determinations (see 74), and the reported differences (without a test of significance) are not large enough to be trustworthy. Also changing the pH of the neurons would have an effect on the receptor membrane potentials, not to mention altering the rates of any enzyme reactions present. The result is thus not at all conclusive.

Two studies, by Sellari Franceschini (121) and by Sellari Franceschini and Camaiori (122), have reported correlations between odor thresholds and thiocyanic acid concentration in the nasal mucosa under various conditions—pregnancy, parturition, and a day's work in a perfume factory. In general the relationship reported is a positive one, although the studies leave much to be desired from the psychophysical point of view, and are, at best, suggestive.

Ghirlanda (55) investigated the hyperosmia which follows injection of

---

[7] It is interesting to note that Angeli and Polverini (14) found a relationship between intensity of odor and oxidizability among the coumarins.

[8] Erb (52) has briefly presented an almost identical theory, without mentioning Niccolini's previous and very extensive work.

insulin to see whether it is accompanied by increased glycogen in nasal mucosa as well as by increased thiocyanic acid concentration. Working on four rabbits, he found no increase in glycogen after injection of insulin, and concluded that the hyperacuity (casual data, apparently) is not linked to a simultaneous modification of glycogen present in the tissues.

Some of the implications of Niccolini's theory are extremely interesting, but lead us so far into the field of ionic organic reactions as to make it impossible to pursue them here. A few comments are, however, in order. Since the important step in stimulation is a shift toward negativity of the receptor field, ionization is important. In some cases, ionization is sufficiently high to permit stimulation without further ado, after the substance has dissolved in the mucus (e.g., ethyl alcohol). In other cases such as benzene, or, especially, a saturated hydrocarbon such as octane, it is evident that ionization is an extremely unlikely possibility. So we must conclude that Niccolini's theories need considerable clarification of detail.

Le Magnen (80, 81) has proposed a humoral theory of olfaction. It is based upon a study of olfactory sensitization by intramuscular injections of an odorous substance. He supposes that the same antiseral physicochemical system which operates in the antigen-antibody reactions reacts with the odorous molecule, because the "odor antibody" has already been established by the penetration of the substance into the organism by some means—hormonal, alimentary, or parenteral. This reaction (type not specified, so that the stimulation problem is not really answered) sets off the nerve impulse. This would explain why certain substances stimulate, but would not explain the problem of qualities without subsidiary assumptions, which are not made, about number of end-organ types or pattern of stimulation, or some similar mechanism for translating stimulation into sensation.

Le Magnen thinks that his theory explains a number of embarrassing facts, such as the recognition of a host by a parasite, or sexual attraction narrowly limited to the species. He also thinks such a theory would explain the great individual differences in olfaction. While it might explain individual differences in sensitivity, it still does not provide us with any theory to explain qualities, and the exact mechanism of olfactory stimulation is tossed in the lap of the immunologist. Nevertheless, such sensitization of living cells to particular substances through exposure is well known in pharmacology and deserves further consideration.

In another article, Le Magnen (78) states that his work (77, 79) and also that of Guillot (56, 57) point clearly to very specific receptors. Le Magnen's theory should thus be considered to involve a large number, as yet undetermined, of specific receptors, with individual differences determined by humoral sensitization, and the mechanism of excitation as yet undetermined.

In sum, we may say that the theories in this group range from fairly simple relationships which have not proved useful to more complicated ones which still require a good deal of research. There is none which can be accepted at the present time as more than suggestive of one possible area for exploration. It is becoming increasingly obvious that more than one mechanism for excitation of the receptors is likely, and theories presenting only one possible relationship are thus exploring only one of several aspects.

F. STERIC THEORIES

There are numerous references to the possibility of a steric factor as an influence of prime importance in olfaction, Ruzicka being among the first to indicate some such possibility, but there are few well-developed or well-documented theories.

Troland (130) has proposed one of the very few theories which attempts to account for all the psychophysical problems, and one of the key concepts is a steric factor. First, he provides for the mechanism of stimulation. Molecules are thought to be adsorbed on the surface of the olfactory hairs or cells, which depolarizes the receptor and thus sets off the nerve impulse. But molecules, to be odorous, must have a form which is so related to the membrane mosaic that the two systems can combine. Olfactory quality is determined primarily by the timing of the nerve impulse, although Troland thinks it possible that the membranes of olfactory cells may be specifically differentiated in a number of different ways. If this is so, then some cells might adsorb one kind (shape) of molecule, and others would adsorb other shapes. Intensity of odor is taken care of by either the total number of similarly excited fibers, or by the total number of impulses in the olfactory nerve. Troland finally suggests that olfactory qualities may be spread out in a kind of unidimensional olfactory spectrum, rather than being classifiable into a definite number of basic types. A similar suggestion has been made by many other workers in the field, who have been impressed by the negative character of the evidence for unique qualities. Troland's theory, although facing squarely the problems involved, leaves some unsolved problems of detail. No evidence is brought to bear on the likelihood of steric facilitation. Evidence is also lacking regarding the determination of quality by the timing of impulses. Timing of impulses throughout the central nervous system is a vastly important problem about which very little is known, and it will probably be some time before a reasonable judgment can be made on this point. Troland's suggestions remain, however, among the most sophisticated in the field.

Pauling (109, 110, 111) has proposed a lock-and-key theory of olfaction, derived from his studies of the specificity of antibodies and enzymes, which is attractive because of the fact that olfactory responses appear to be of approximately the same order of specificity as antibody and enzyme response. He states:

> The specificity of the catalytic activity of enzymes is due to a surface configuration of the enzyme such as to make the enzyme complementary to the substrate molecule, or, rather, to the substrate molecule in the strained state that occurs during the catalyzed reaction. The specific action of drugs and bactericidal substances has a similar explanation. Even the senses of taste and odor are based upon molecular configuration rather than upon ordinary chemical properties—a molecule which has the same shape as a camphor molecule will smell like camphor even though it may be quite unrelated to camphor chemically (110, p. 1065).

The theory is of course incomplete: there is no mention of the kinds of reaction which are facilitated by the greater proximity of important parts of odorous molecules to specific areas of the receptors, and no discussion of the problem of qualities, but the theory does offer an explanation of why chemically different substances may smell the same, and why chemically similar substances may smell different. The further advantage, mentioned above, of allowing for the great specificity of response which has thus far been the major positive result of olfactory research is not inconsiderable. So far there is no compelling evidence that the construction of two molecules of very similar shape will result in compounds smelling alike, but as one factor in facilitating some kind of response between the odorous substance and receptor it should not be dismissed. This theory is well described by Haagen-Smit (59).

It is interesting that a similar suggestion is made by Dethier and Chadwick, arising from their investigations on contact chemoreception in insects. They state: "In as much as chain length appears to be of importance as far as stimulation is concerned, it is not improbable that an explanation of the differences between straight diols and polyethylene or polypropylene compounds may be associated with the difference in molecular shape" (40, p. 128).

Moncrieff (92, 93) later proposed a very similar theory which he calls a "configurational" theory. No evidence is adduced for it, only arguments. He also assumes a number of specialized receptors, the number (between

6 and 12) being arrived at by the following process: "Probably, in view of the number, which is between 20 and 30, of fibers in the human olfactory nerve, and taking into account Nature's usual provision of a 'spare' in case of accident (or possibly dual provision for time and distance stereo-effects) there are between six and 12 types of olfactory receptors" (92, p. 282). Unfortunately, the number of fibers assumed is very seriously in error [Adrian (2) would have it 10 to 100 million fibers! Cf. also 12], and there is no guarantee that "Nature" always provides a spare. The nature of the energy change within the receptor is also subject to theorizing by Moncrieff: he thinks it is probably caused by a reduction in surface tension. As we have previously pointed out, there are objections to this suggestion (cf. Watson).

Steric factors are sometimes used as accessory mechanisms in theories in which the basic mechanism is otherwise classified. Ehrensvärd and Cheesman (49) consider steric factors to play a part in the alteration of the interface potential (vide infra). And recent work by Ruzicka and his associates (114, 115) has pointed up the importance of the spatial structure of the molecule in odor: "Der räumliche Bau der Molekule spielt dabei eine wichtige Rolle. Es ist bemerkenswert, dass der Geruch in gleichem Sinne von der Konfiguration beinflusst wird wie die männliche Hormonwirkung" (114, p. 619).

Guillot (58) has entered the lists on the side of the molecular configurationalists. He is of the opinion that chemically different substances give rise to similar odors only when the general molecular configuration is the same. There must be, he says, a direct relation between each type of molecular structure and a particular type of receptor, possibly a "compatibility" of configuration between the odorous molecule and one of the essential components of the sensory cell. The physiological process is envisaged as similar to those involved in the action of vitamins, antibodies, etc. Substances with complex odors are thought to be molecules which are either easily deformed by several different receptors in the process of adsorption, or adsorbed by different receptors at different molecular regions. Upon the basis of his work on partial anosmias (56, 57) Guillot posits a large number of fundamental odors, in line with the great variety of molecular structures which occurs among odorants.

Angeli (13) suggests that steric hindrance may be a factor in odor, and that the deformation of a ring, for example, may be more important than the specific substituent in determining odor. Although this theory is not a complete one, it is a piece with other steric theories, and deserves similar consideration.

Nerdel and Spaeth (99) support cautiously a theory combining steric factors and osmophoric groups. They think that probably the surface form of the molecule determines the kind and place of adsorption, but that the (usually polar) osmophore groups either participate in reactions with substances in the receptor or form preliminary stages to reactions. But then they state that "purer" odors result as the molecular structure becomes more important, and that the reactions of the osmophore groups result in "other nasal sensations"—pain and cold. Since these are not olfactory sensations, odor is left dependent upon the shape of the molecule.

In concluding the discussion of steric theories, some comments should be made. The importance of the configuration of molecules is recognized in immunology (110, 111), and in pharmacology (8, 35, 36, 37). The shapes of molecules, or the shapes of critical regions of molecules, are thought to be of importance in determining their biological effects. But in order to understand why steric factors should be important, it is necessary to consider how they affect reactions. The shape of a molecule is of importance in relationship to the shapes of other molecules. If the reactive loci of two complex molecules cannot approach each other sufficiently closely, no reaction can occur, and we speak of "steric hindrance." Thus, the drug—or odor—must be a "key" which fits the biological "lock." The "reaction" may involve anything from the weakest van der Waals bond, to the formation of new covalent bonds. From the above very brief statements, it is apparent that steric factors in and of themselves do not exhaust the problem of how odor and receptor interact. We still need decide where and how the odor and the receptor cell react—whether by "mass action," adsorption on the cell surface, or some other means. Therefore, although knowledge of how molecular configuration affects odor is of considerable practical importance, it is not completely satisfying to leave the problem there.[9]

G. Phase Boundary Theories

Phase boundary theories are primarily concerned with the mechanism of stimulation of the receptor, although they may secondarily be concerned with the problem of quality. Kremer (76) developed very early an olfactory theory dependent upon an assumed similarity between narcotics and odor stimuli: that both require the passage of volatile material from an aqueous phase, where it is dissolved, to a colloidal system. He believed that odorous substances were first dissolved in the mucus, then either absorbed into the olfactory cells by the phosphatides, probably lecithin, or adsorbed onto the surface. In either event the cellular balance would be disturbed, and the nerve impulse initiated. No explanation of olfactory qualities is offered.

---

[9] Change in steric properties is not the only way of explaining the influence of different molecular configurations or components on biological activity. For a full discussion, see Albert (8).

Ehrensvärd (48, 49, 50, 51) places the problem of olfaction in the same general category as any of the problems involving the influence of organic compounds on physiological systems. His research has been directed to a highly simplified *in vitro* analogy of the olfactory system, the point of interest for Ehrensvärd being the potentials arising at the interfaces between nonmiscible solutions. The analogy with the olfactory system is the lipoid cellular surface versus the aqueous mucus or between water and lipoid phases within the membrane or the cell itself. His results are very interesting and examples are given here because they will serve to illustrate a type of result generally obtained in this area. In the system 1-n-KCl/A/x-n-KCl, where A is $ClC_6H_4NH_2$ he finds rather remarkable differences in the potentials at the interface when the Cl atom in the amines is shifted in position. With the Cl atom in the ortho position, the potential is about 55 mv, in the meta position about 75 mv, and in the para position about 80 mv. Even greater differences appear in some of the other systems investigated. A series of various aliphatic compounds gave potential differences ranging —140 to 220 mv. Although there is no proof that such changes are the effective ones in olfaction, it is nevertheless instructive to see that measurable energy changes do occur in somewhat analogous systems, under the influence of changes which we know effect changes in odor; change in concentration in some substances, and change in position of some atom or group, and that these changes are of the proper order of magnitude.

On the basis of this work Ehrensvärd has developed a theory of olfaction. Analysis of odorous material by the olfactory receptors, he says, is governed by not one but several structural factors. The more limited the possibilities for direct chemical interaction of odorous molecules with molecules in the receptor, the greater are the number of odor-variants, because in this case small changes in energy relationships within the molecule are effective. The "analysis" of a compound by the receptor is directed at the molecular region of greatest electron or proton motion and the less distinctive this motion is the greater the importance of the spatial structure of the molecule. (The term "proton motion" is Ehrensvärd's but has little meaning.) In other words, in those cases where direct chemical reaction does occur between odorous substance and receptor substance, there should be only a few kinds of odors which are not modified by slight modifications of the odorous molecule. On the other hand, where the receptors are stimulated by some effect such as potentials arising at an interface, there should occur a virtually infinite variety of odors, and a small change in the odorous molecule, such as occurs in the d- and l-forms of a substance, may be sufficient to cause a change in odor quality. For Ehrensvärd there is no "monosmic" olfactory stimulus, and no continuum of qualities.

This theory is concerned primarily with the effective stimulus, or stimuli, and a dual theory is proposed, both direct chemical interaction and alteration of potential in the cell by means of potentials arising at some interface. Many modern theorists are coming to the conclusion that a single mechanism is not sufficient, and it is becoming increasingly likely that direct chemical interaction does occur under some circumstances (see the general discussion). The second mechanism here proposed remains a possibility. Dethier and Charwick (38, 39) have thought that a similar mechanism must be part of the mechanism of contact chemoreception in insects: "Whatever the exact mechanism of their action, the majority of data in the now rather extensive literature show that the physiological activity of the alcohols (and of many other compounds) runs parallel with the rate of their adsorption at a lipid-water interface or with their distribution between lipid and water" (38, p. 251). This may govern the rate of access to the system rather than the final interaction, however.

The problem of quality is disposed of by assigning it to the variation of potential at the oil-water interface. Unfortunately this reckons without the all-or-nothing law. The variation must be in the final nerve impulse, and while this is still a possibility, due to the variation allowed by the relative refractory period, it is not nearly as clear-cut as it seems. Either special receptors must be assumed, with different thresholds, or some resort must be made to timing of impulses (cf. Troland) within the single nerve fiber or sub-systems. Ehrensvärd and Cheesman (49) also suggest an accessory steric factor in the influence of foreign substances upon the potentials at an interface. Finally, it must be pointed out that the emphasis on an oil-water interface ignores the probability of adsorbed protein on the receptor surface, thus oversimplifying the actual situation (35).

Middleton, apparently unaware of Ehrensvärd's monograph, has recently suggested that the electric charge at interfaces should be investigated *in vitro*, using a synthetic emulsion system and simple homologous series (88).

Austerweil (15) has criticized Ehrensvärd's theory on the ground that his work involved liquid/liquid interfaces, whereas, in Austerweil's opinion, the olfactory system involves solid/liquid interfaces. The olfactory system undoubtedly has both, and it is not possible to accept or reject one or the other.[10] Austerweil's own theory is that the important factor in stimulation is the changing of the interface tension between the olfactory hairs and the liquids that bathe them when one or the other adsorbs or dissolves the odorous molecules.

McCord and Witheridge (84) support what they call an electrochemical theory. They lay emphasis upon the disturbance of the electrostatic field at the moment of solution of an odorous molecule in the mucus, due to modification of bonding angles. Different amounts of potential difference thus provided are thought to account for different odors. How variations in potential difference are translated into qualities is not clear. It is difficult to see how mere gradations of net change in potential difference could be made to account for all phenomena of stimulation and of quality without subsidiary factors.

Bungenberg de Jong and Saubert (28, 29) have proposed a model for the stimulation of the sense of smell based upon *in vitro* experiments with coacervates (aggregations of molecules in a liquid medium). They have been impressed by a number of similarities between the responses of phosphatide and oleate coacervates and the olfactory system, notably the great sensitivity to organic nonelectrolytes in small concentrations. They find that the phosphatide and oleate coacervates are highly sensitive to the vapors of many organic compounds, and that the changes are reversible. These changes are described as an increase or decrease in the average distance between the phosphatide or oleate molecules in the coacervate. Some of the effects described are interesting. Aromatic hydrocarbons have a strong shrinking effect (decrease of average distance between molecules); aliphatic hydrocarbons have a weak shrinking effect, unless there is a halogen substitution, in which event this effect is greatly strengthened; but paraldehyde, acetone, and ethyl alcohol have the opposite effect: they expand the coacervate. The authors find that with certain substances a low concentration gives rise to the shrinking effect, whereas a high concentration gives rise to the expansion effect. In homologous series of alcohols and ketones one finds a reversal of effect at some specific point, e.g., methyl, ethyl, and n-propyl alcohols cause expansion, but n-butyl, n-amyl, n-hexyl, and n-heptyl cause shrinking. Further, the effect of alcohols depends upon the branching of the carbon chain: n-butyl alcohol is a shrinking agent, but tertiary butyl alcohol causes expansion. Their theory of action of the organic nonelectrolytes is as follows: hydrocarbons work, by means of van der Waals forces, as cohesion agents for the hydrocarbon chains of the oleate molecules in the coacervate and thereby cause shrinking. But the introduction of hydrophilic groups ($OH_2 = O$) in the hydrocarbon molecule causes the opposite response. Where the reversal of action obtains in an homologous series, it is because one effect has at some specific point been overbalanced by the other.

---

[10] Since it is remotely possible that the olfactory receptors may at times project through the mucus (73), perhaps a membrane-air interface needs consideration. Although such a possibility would help explain the lack of odor stimulation from substances in aqueous solution (61), or the odor of carbon aerosol (133), it rests on such tentative grounds at present that it will be ignored pending further evidence. But even if hairs are found to project from the mucus, it is unlikely that they would be without a film of adsorbed water.

The authors feel that such systems can serve as models for the olfactory system because: (*a*) the coacervates are sensitive to almost all sufficiently volatile substances, (*b*) the changes are reversible (which is a great advantage in attempting to explain why adaptation does not occur immediately in olfaction), (*c*) the change is of such a nature that in analogous systems which are much poorer in water they would be manifest as permeability changes, which might be expected to initiate the nerve impulse, (*d*) the stimulating substance reaches the "receptor" from a gas phase, through a watery fluid, and in both the olfactory and coacervate systems a stimulus must have both sufficient volatility and sufficient water solubility before its other attributes are of any moment.

The method of actual stimulation is thus conceived to be a change in the average distance of the very heavy molecules in the receptor cell, which increases or decreases its permeability. This implies a dual action within the single cell, which one might suppose to be followed by an increase or decrease in the number of nerve impulses. The authors do not actually make this proposal, although they imply it when they seek, unsuccessfully, to correlate expansion and contraction with Zwaardemaker's two general classes of odors: nutritive odors and odors of decomposition. There is at present no evidence of a "steady state" of olfactory nerve impulses which can be modulated in either direction.

As to olfactory qualities, the authors propose "at least one kind of cell" for each of Zwaardemaker's nine classes of odors, each one, we presume, containing a different sort of substrate which would be differently affected by various odorous substances. It is suggested that anisotropic systems may play a part. It is unfortunate that modern authors see fit to base their hypotheses upon Zwaardemaker's classification. There is no odor classification which has stood the test of research. The only dependable modern studies of classes of odors are those of Guillot (56, 57, 58) and Le Magnen (77), both as yet incomplete, and the indications from these is that the results will not be simple.

This theory is closely related to that of Ehrensvärd (48). The essential difference lies in the aspect chosen for measurement—electrical potential, or molecular rearrangement.

H. Enzyme Theories

Among the most modern approaches to the problem of olfactory stimulation are the enzyme theories, in which the primary effect of an odorous substance is thought to be on the enzymes of the olfactory cell. Kistiakowsky (75), a chemist, believes that the basic principle underlying olfaction is an enzyme reaction, that the inhibition of certain enzymes contained in the olfactory cells is responsible for odor. A system of enzymatically catalyzed reactions, represented by $A \longrightarrow A' \longrightarrow A''$, $B \longrightarrow B' \longrightarrow B''$, etc., is causally related to signals in the olfactory nerve. Each step is catalyzed by a separate enzyme and each of the compounds (A′, B′, C′) is capable of causing a signal in a particular class of nerve fibers, when its concentration is altered. The effect of an odorous compound is the inhibition of one or more of the enzymes, causing a shift in the relative concentrations of A′, B′, C′, which sets off the nerve impulses. The number of separate enzyme systems (A, B, C) is thought to correspond to the number of basic smells.[11]

This theory, which its author states is pure speculation, is said to account for: (*a*) the extreme sensitivity of the sense of smell, because the quantities of enzymes involved may be very minute, (*b*) the wide range of compounds eliciting odor, because enzymes are frequently inhibited by a great variety of compounds and yet show definite selectivity, (*c*) complex odors giving apparent changes in quality with concentration, because complex odors are the result of the inhibition of several enzyme systems, and with increasing concentration some systems may be almost totally inhibited, (*d*) the reversibility (?) of most odors, because many enzyme inhibitions are reversible, although there also exists the possibility of nonreversible inhibition, which corresponds to the persistence of certain odors (relatively long adaptation time?). This theory is very general, but attempts to account for all phases of the theoretical problem: (*a*) the mechanism of receptor activation (enzyme inhibition), (*b*) the qualitative differences among odors (a number of different enzyme systems and combinations of these), and (*c*) the mechanism of representation of intensity (the amount of inhibition of a particular enzyme system).

---

[11] There is no real reason for assuming that the "basic" smells, or unique olfactory qualities, if such are ever discovered, will correspond to the physiological primaries. Such is almost certainly not the case in vision. Guillot (57) has found two separate and distinct anosmias for musk-like odors, for example, which would lead one to suspect that introspective techniques are going to be of little use in olfactory classification, as indeed they have been until now.

Alexander (11) has criticized this theory as an emasculated form of his own theory. Actually Kistiakowsky's theory is much more thoroughly worked out, although the specific points do follow very easily from enzyme chemistry, and there seems to be no reason to doubt its originality. Alexander's original state of the theory (9) consists of a brief statement in a textbook on colloid chemistry, where no one interested in olfaction would be likely to find it. The theory merely states that odorous substances may act as modifiers of cellular catalysts. "Modification" is elucidated only in very general terms. A more recent publication(10) adds nothing specific.

The most complete statement of the enzyme theory is given by Baradi and Bourne (17), who have followed up Kistiakowsky's hypothesis with some experimental results. Bourne (24) found that alkaline phosphatase is present in relatively high concentrations in the taste buds of several species, or in the epithelium overlying them, and it is also present in appreciable concentrations in the olfactory mucosa. The experimental results of Baradi and Bourne deal with the gustatory mechanism, where sapid substances are shown to inhibit two enzymes. The argument for such a mechanism for odor is purely analogical, and taste and smell cannot be considered closely similar. There are enough suggestions of similarity to make worth while an experimental test of the inhibition of olfactory cell enzymes, but enough differences to render the assumption of identity tenuous. A later publication of Baradi and Bourne (18) presents more detailed results and a more complete theory of gustation, and suggests that olfaction may operate in the same way: a large number of enzymes, which may exist in a number of different sites (six for gustation) and which may have their activity accelerated, inhibited, or unchanged by an odorous substance. Such changes give rise to ionic changes which set off the nerve impulse. Each odorous substance is presumed to give rise to a distinct pattern of nerve impulses from the whole olfactory receptor region. Thus the problem of quality is one of stimulus pattern.

In considering enzyme theories, it is well to keep in mind that all living activity depends on enzymes—indeed, this is a truism. It is now well known that many drug actions depend upon an alteration of enzyme activity and it may be (cf. Hensel and Zotterman, 68) that the arousal of cold receptors involves an enzyme. It is not feasible to review here all of the ways in which a foreign substance can inhibit or facilitate an enzyme. These may be culled readily from Danielli (36), Albert (8), Höber (72), and Rideal

(118). An interesting discussion of the way in which enzyme inhibition may lead to facilitation may be found in Woolley (136, pp. 111-112). In sum, enzyme theory is a most promising and sophisticated approach to olfaction.

I. Concluding Comments

It is apparent that no final choice among extant theories can be made on the basis of present knowledge, nor is any of them more than merely plausible. If we cast olfaction into the same general category as pharmacology, and it is clearly reasonable that we should since both involve the action of foreign molecules on living cells, it is possible to suggest hypotheses concerning the problem of olfactory stimulation. It is hoped that the following comments may serve to clarify the issues.

It will be remembered that the olfactory receptors are primitive in that they are really afferent neurons without specialized epithelial accessory structures. In considering how they act, the question comes to mind as to whether they are normally, that is in the resting state, quiescent or active. Are they more like the bulk of visual receptors, for example, than they are like some of the proprioceptors? Fortunately, most of the further discussion does not require a decision on this point, since accounting for *activation* is not greatly different from accounting for *modulation,* although in the latter case there is an extra degree of freedom—that is, activity may be reduced as well as enhanced. If Adrian's most recent results turn out to be correct, the receptors are normally quiescent (3). [Hasama (62) does not agree and it is difficult to evaluate the effect of the anesthesia.] We shall continue as though this were so, and only minor alterations in the discussion will be necessary if this assumption should prove incorrect.

Since alterations in the membrane potential of the soma of a neuron can give rise to repetitive discharges in the associated axon, it seems reasonable to assume that this is the mode of stimulation of the olfactory receptors. Whether this activation takes place in the olfactory hairs, or in the cell-body itself is a matter for conjecture. Basic theory would not be altered very much one way or the other, except for one interesting possibility. This possibility is that odors sensitize the receptors by altering their membrane potentials, but that mechanical stimulation of the hairs is necessary to produce nerve impulses. Observations that perfectly still odorous air is not sensed lend support to this speculation (61), but recent research on intravenously injected odors tends to rule it out (21). In any event, the basic problem is to account for the activating or sensitizing shifts in membrane potential.

The first problem is that of locus of action. Must an odor dissolve in the cell substance to stimulate as Cohn (38) believed, or is the action on the cell surface? Drugs can work in either way, with surface activity probably more usual (35). However, some substances, at least, must reach the cell interior to affect its activity (36) and it is also possible that the essential reactions take place within the membrane itself.[12] A final decision must await more direct evidence than is now available, but on the basis of the analogy with drug action, it may tentatively be concluded that *all* these loci may be important, depending upon the particular odor in question, although surface activity is probably the most common.

The second problem concerns the nature of the interaction between odor and receptor. We have reviewed a large number of suggestions as to what this interaction might be, and have tried to give the strengths (if any) and weaknesses of each. At this point it might be well to indicate very briefly indeed which ideas we feel deserve the greatest attention. Actually, there are several, any one or more of which may be important for a particular odor—there is no necessity for assuming that all odors act in the same way. These ideas are listed below:

1. Alteration of interface potentials, either at the cell surface, or between phases within the cell. Such alterations could take place in several ways: (*a*) shift of mucus pH (cf. Niccolini, 107); (*b*) tying up or destroying the polar groups (—OH or —SH, most likely) of the membrane surface, as by trivalent arsenic; (*c*) adsorption at the interface (cf. the discussion of steric theories); (*d*) adsorption at the air-mucus interface; (*e*) alteration in salt content of the mucus; and (*f*) shifts in the spreading properties of the cell proteins. The work on intravenous odors would make it unlikely, of course, that any of these possibilities is the sole mechanism.

2. Absorption: mass effects in membrane or cytoplasm. Accumulation of inert foreign molecules in a cell may profoundly alter its activities. Benzene or octane come to mind as likely to act in this way.

3. Absorption: denaturing or reorganizing effects. It is possible that after absorption, chemical reaction may take place. For example, it would seem unlikely that trivalent arsenic, halogens, aldehydes, alkylating agents, and oxidizing agents would not react with the thiol groups in the cell proteins.

4. Enzyme inhibition or activation. This suggestion has been discussed

---

[12] It should be recognized that membrane potentials may actually exist between layers *within* the membrane rather than between membrane and extra-cellular medium (cf. Lorente de Nó, 86).

References

1. Adrian, E. D. Olfactory reactions in the brain of the hedgehog. *J. Physiol.,* 1942, **100**, 459-473.
2. ———. The sense of smell. *Advanc. Sci.,* 1948, **4**, 287-292.
3. ———. Olfactory Discrimination. Lecture at Calif. Inst. of Technol., Oct. 18, 1949.
4. ———. Sensory Integration. Univ. Press of Liverpool, 1949. Pp. 20.
5. ———. The electrical activity of the mammalian olfactory bulb. *EEG & Clin. Neurophysiol.,* 1950, **2**, 377-388.
6. ———. Olfactory discrimination. *L'Année Psychol.,* 1951, **50**, 107-113.
7. Airkem. Odors and the Sense of Smell. New York: Airkem, 1952. Pp. 342.
8. Albert, A. Selective Toxicity. London: Methuen, 1951. Pp. xii+228.
9. Alexander, J. Colloid Chemistry. (4th Ed.). New York: Van Nostrand, 1937. (Pp. 177, 179, 376.)
10. ———. Odor detection. *Chem. & Eng. News,* 1949, **27**, 2227.
11. ———. Catalysis and olfaction. *Science,* 1951, **114**, 674-675.
12. Allison, A. C., & Warwick, R. T. T. Quantitative observations of the olfactory system of the rabbit. *Brain,* 1949, **72**, 186-197.
13. Angeli, A. Sopra alcune relazioni fra costituzione ed odore. *Gazz. Chim. Ital.,* 1930, **60**, 939-946.
14. Angeli, A., & Polverini, A. Sopra alcune relazioni fra odore e comportamento chimico. *Gazz. Chim. Ital.,* 1931, **61**, 276-280.
15. Austerweil, G. Sur la théorie de l'olfaction d'après Ehrensvärd. *Parfumerie,* 1943, **1**, 217-218.
16. Banerji, M. N. Incidence of smell: Theory of surface friction. *Indian J. Psychol.,* 1930, **5**, 87-94.
17. Baradi, A. F., & Bourne, G. H. Localization of gustatory and olfactory enzymes in the rabbit, and the problems of taste and smell. *Nature* (Lond.), 1951, **168**, 977-979.
18. ———. Theory of tastes and odors. *Science,* 1951, **113**, 660-661.
19. Beck, L. H. Osmics: Olfaction. Medical Physics (O. Glasser, *Ed.*) Vol. 2, Chicago: Yearbook Publishers, 1950 (pp. 659-664).
20. Beck, L. H., & Miles, W. R. Some theoretical and experimental relationships between infrared absorption and olfaction. *Science,* 1947, **106**, 511. (Abstract.)
21. Bednár, M., & Langfelder, O. Über das intravenöse (hämatogene) Riechen. *Monatsschr. f. Ohrenheilk. u. Laryngo-Rhinol.,* 1930, **64**, 1133-1139.
22. Börnstein, W. S. Cortical representation of taste in man and monkey; functional and anatomical relations of taste, olfaction, and somatic sensibility. *Yale J. Biol. Med.,* 1939-40, **12**, 719-736.
23. Boring, E. G. Sensation and Perception in the History of Experimental Psychology. New York: Appleton-Century, 1942. (Pp. 437-462.)
24. Bourne, G. H. Alkaline phosphatase in taste buds and nasal mucosa. *Nature* (Lond.), 1948, **161**, 445-446.
25. v. Braun, J., & Anton, E. Geruch und molekulare Asymmetrie: IV. Mitteilung. *Ber. d. Deut. Chem. Gesell.,* 1927, **60**, 2438-2446.
26. v. Braun, J., & Kaiser, W. Geruch und molekulare Asymmetrie. *Ber. d. Deut. Chem. Gesell.,* 1923, **56**, (2), 2268-2274.
27. Brink, F., Jr. Excitation and conduction in the neuron. Chap. 2 in: Stevens, S. S. (*Ed.*), *Handbook of Experimental Psychology.* New York: Wiley, 1951. (Pp. 50-93.)

28. BUNGENBERG DE JONG, H. G., & SAUBERT, G. G. P. Kolloidmodelle biologischer Vorgänge. Für Dämpfe organischer nichtelektrolyte empfindliche Koazervate als Modelle der Geruchssinnreizung. *Protoplasma,* 1937, **28**, 329-344.

29. ———. Models for the stimulation of the organ of smell. *Kon. Akad. van Weten. Amster., Proc. Sec. Sci.,* 1937, **40**, 302-306.

30. CHADWICK, L. E., & DETHIER, V. G. The relationship between chemical structure and the response of blowflies to tarsal stimulation by aliphatic acids. *J. Gen. Physiol.,* 1946-47, **30**, 255-262.

31. ———. Stimulation of tarsal receptors of the blowfly by aliphatic aldehydes and ketones. *J. Gen. Physiol.,* 1949, **32**, 445-452.

32. CLARK, W. E. L. Projection of the olfactory epithelium on to the olfactory bulb: a correction. *Nature* (Lond.), 1950, **165**, 452-453.

33. CLARK, W. E. L., & WARWICK, R. T. T. The pattern of olfactory innervation. *J. Neur., Neurosurg., & Psych.,* 1946, **9**, 101-111.

34. COHN, G. Die Riechstoffe. (Second Ed.) Braunschweig: Verlag Vieweg, 1924. Pp. viii+216.

35. DANIELLI, J. F. The cell surface and cell physiology. Chap. 4 in Bourne, G. (*Ed.*), *Cytology and Cell Physiology.* (Second Ed.) Oxford: Clarendon Press, 1951. (Pp. 150-182.)

36. DANIELLI, J. F. Cell physiology and pharmacology. New York: Elsevier, 1950. Pp. viii+156.

37. DANIELLI, J. F., & DAVIES, J. T. Reactions at interfaces in relation to biological problems. In Nord, F. F. (*Ed.*) *Advances in Enzymology.* New York: Interscience, 1951. (Pp. 35-89).

38. DETHIER, V. G., & CHADWICK, L. E. Rejection thresholds of the blowfly for a series of aliphatic alcohols. *J. Gen. Physiol.,* 1946-47, **30**, 247-253.

39. ———. Chemoreception in insects. *Physiol. Rev.,* 1948, **28**, 220-254.

40. ———. The stimulating effect of glycols and their polymers on the tarsal receptors of blowflies. *J. Gen. Physiol.,* 1948, **32**, 139-151.

41. DOLL, W., & BOURNOT, K. Über den Geruch optischer Antipoden, *Pharmazie,* 1949, **4**, 224-227.

42. DURRANS, T. H. Relation of odour to chemical constitution. *Perf. Ess. Oil Rec.,* 1919, **10**, 266-267.

43. ———. The relationship between odour and chemical constitution. *Perf. Ess. Oil Rec.,* 1919, **10**, 104-136.

44. ———. The "residual affinity" odour theory. *Perf. Ess. Oil Rec.,* 1920, **11**, 391-393.

45. DYSON, G. M. Some aspects of the vibration theory of odour. *Perf. Ess. Oil Rec.,* 1928, **19**, 456-459.

46. ———. Raman effect and the concept of odour. *Perf. Ess. Oil Rec.,* 1937, **28**, 13-19.

47. ———. The scientific basis of odour. *Chem. & Ind.,* 1938, **57**, 647-651.

48. EHRENSVÄRD, G. Über die Primärvorgänge bei Chemozeptorenbeeinflussung. *Acta Physiol. Scand.,* 1942, **3**, Suppl. IX. Pp. 151.

49. EHRENSVÄRD, G., & CHEESMAN, D. F. On the specific adsorption of alcohols at the salicyl-aldehyde/water intersurface. *Science,* 1941, **94**, 23-25.

50. EHRENSVÄRD, G., & SILLÉN, L. G. Adsorption potentials and aging liquids. *Nature* (Lond.), 1938, **142**, 396-397.

51. ———. Adsorption potentials at polar liquid-liquid interfaces. *Nature* (Lond.), 1938, **141**, 788-789.

52. ERB, R. C. Electrons as a cause of odor. *Amer. Perfumer & Ess. Oil Rev.*, 1947, **50**, 561.

53. FARADAY SOCIETY. Modes of Drug Action. London: Gurney & Jackson, 1943.

54. GAVAUDAN, P., POUSSEL, H., BREBION, G., & SCHUTZENBERGER, M. P. L'étude des conditions thermodynamiques de l'excitation olfactive et les théories de l'olfaction. *C. R. Acad. Sci.*, 1948, **226**, 1395-1396.

55. GHIRLANDA, M. Sulla presenza di glicogena nella mucosa olfattorio. Puo'avere il glicogena nasale rapporto con la funzione dell'olfatto? *Atti Accad. Fisiocrit. Siena.*, 1950, **18**, 407-412.

56. GUILLOT, M. Anosmies partielles et odeurs fondamentales. *C. R. Acad. Sci.*, 1948, **226**, 1307-1309.

57. ———. Sur quelques caractères des phénomènes d'anosmie partielle. *C. R. Soc. Biol.*, Paris, 1948, **142**, 161-162.

58. ———. Sur la rélation entre l'odeur et la structure moléculaire. *C. R. Acad. Sci.*, 1948, **226**, 1472-1474.

59. HAAGEN-SMIT, A. J. Smell and taste. *Sci. Amer.*, 1952, **186**, 28-32.

60. HAMELIN, G. Flavors, odors, and infra-red rays. *Sci. Amer. Mo.*, 1921, **3**, 519-522.

61. HARTRIDGE, H. Note on the sense of smell. *J. Physiol.*, 1920-21, **54**, xxxix-xli.

62. HASAMA, B. Über die elektrischen Begleiterscheinungen an der Riechsphäre bei der Geruchsempfindung. *Pfl. Arch. ges. Physiol.*, 1934, **234**, 748-755.

63. HELLER, H. Über die Geruchstheorie von Teudt. *Biol. Zentralbl.*, 1919, **39**, 364-370.

64. ———. A new theory of odor and scent. *Amer. Perfumer*, 1919-20, **14**, 365-366.

65. ———. Teudt's theory. *Perf. Ess. Oil Rec.*, 1920, **11**, 38-39.

66. ———. Über die Geruchstheorie von Teudt: II. *Biol. Zentralbl.*, 1921, **41**, 138-142.

67. HENNING, H. Der Geruch (2nd Ed.). Leipzig: Barth, 1924. Pp. 434.

68. HENSEL, H., & ZOTTERMAN, Y. The effect of menthol on the thermoreceptors. *Acta. Physiol. Scand.*, 1951, **24**, 27-34.

69. HESSE, W. Zur Pathologie des Geruchssinnes. *Zsch. ges. Inn. Med.*, 1950, **5**, 635-637.

70. HEYNINX, M. La physiologie de l'olfaction. *Rev. Oto-Neuro-Ophthal.*, 1933, **11**, 10-19.

71. HIBBEN, J. H. The Raman Effect and Its Chemical Applications. New York: Reinhold, 1939. Pp. 544.

72. HÖBER, R. Physical Chemistry of Cells and Tissues. Philadelphia: Blakiston, 1945. Pp. xiii+676.

73. HOPKINS, A. E. The olfactory receptors in vertebrates. *J. Comp. Neurol.*, 1926, **41**, 253-289.

74. JONES, F. N. A test of the validity of the Elsberg method of olfactometry. *Amer. J. Psychol.*, 1953, **66**, 81-85.

75. KISTIAKOWSKY, G. B. On the theory of odors. *Science*, 1950, **112**, 154-155.

76. KREMER, J. H. Adsorption de matières odorantes et de narcotiques odorants par les lipoïdes. *Arch. Néerland. de Physiol.*, 1916-17, **1**, 715-725.

77. LEMAGNEN, J. Étude d'une méthode d'analyse qualitative de l'olfaction. *L'Année Psychol.*, 1942-43, **43-44**, 249-264. (Publication date: 1947.)

78. ———. Travaux récents sur les sensibilités chimiques. *L'Année Psychol.*, 1946-47, **47-48**, 247-254. (Publication date: 1949.)

79. ———. Analyse d'odeurs complexes et homologues par fatigue. *C. R. Acad. Sci.*, 1948, **226**, 753-754.

80. ———. Étude d'un phénomène de sensibilisation olfactive. *C. R. Acad. Sci.*, 1949, **228**, 122-124.

81. ———. Variations spécifiques des seuils olfactifs chez l'homme sous actions androgène et oestrogène. *C. R. Acad. Sci.*, 1949, **228**, 947-948.

82. LEMAGNEN, J., & RAPAPORT, A. Essai de détermination du rôle de la vitamine A dans le méchanisme de l'olfaction chez le rat blanc. *C. R. Soc. Biol.*, Paris, 1951, **145**, 800-803.

83. LORENTE, DE NÓ, R. A study of nerve physiology. Studies from Rockefeller Institute for Medical Research, 1947, No. 131, pp. 496; No. 132, pp. 497.

84. MCCORD, C. P., & WITHERIDGE, W. N. Odors: Physiology and Control. New York: McGraw-Hill, 1949. (Pp. 22-25.)

85. MANCI, F. Modificazioni della acuità olfattoria determinata per mezzo di spostamenti dell'equilibrio ionico del muco nasale. *Riv. Clin. Med.* (suppl. II, *Marg. Otolaryng.*, vol. 5, F. II), 1946, **46**, 367-378.

86. MASAJI, S. Vitalfärbung des Riechepithels der Maus mit Trypanblau. *Zsch. f. Zellforsch. u. Mikroskop. Anat., Abt. A: Allg. Zellforsch. u. Mikroskop. Anat.*, 1941, **31**, 218-223.

87. MAXIMOW, A. A., & BLOOM, W. A Textbook of Histology. (5th Ed.) Philadelphia: Saunders, 1948. (Pp. 453-456.)

88. MIDDLETON, J. W. Present theories of olfaction—a critical review. *Proc. Sci. Sec., Toilet Goods Assoc.*, 1950, No. 14, 4-9.

89. MILES, W. R., & BECK, L. H. Infra-red absorption in field studies of olfaction in bees. *Science*, 1947, **106**, 512. (Abstract.)

90. ———. Infra-red absorption in field studies of olfaction in honeybees. *Proc. Nat. Acad. Sci.*, 1949, **35**, 292-310.

91. MONCRIEFF, R. W. The Chemical Senses. London: Leonard Hill, 1944: New York: Wiley, 1946. Pp. vii+424.

92. ———. A new theory of odour. *Perf. Ess. Oil Rec.*, 1949, **40**, 279-285.

93. ———. The origin of odour. *Internat. Perfumer*, 1950, No. 1, 13-21.

94. MORGAN, C. T., & STELLAR, E. Physiological Psychology. (Second ed.) New York: McGraw-Hill, 1950. (Pp. 107-122.)

95. MORIMURA, S. Untersuchung über den Geruchssinn. *Tokohu J. Exper. Med.*, 1933-34, **22**, 417-448.

96. MÜLLER, A. A dipolar theory of the sense of odour. *Perf. Ess. Oil Rec.*, 1936, **27**, 202-205.

97. ———. Die Luminescenz (Fluorescenz) der Riechstoffe im filtrierten Ultravioletten Licht. *J. f. prakt. Chem.*, 1939-40, **154**, 209-218.

98. NATIONAL BUREAU OF STANDARDS. American Petroleum Institute Research Project 44, Nos. 104, 147, 242, 698.

99. NERDEL, F., & SPAETH, I. Geruch und Konstitution. *Angew. Chem.*, 1951, **63**, 545-550.

100. NICCOLINI, P. Contributo allo studio di alcune sostanze odorose: la cumarina sintetica ed alcuni suoi derivati. *Arch. Ital. Sci. Farmacol.*, 1932, **1**, 253-285.

101. ———. Contributo allo studio fisio-farmacologico dell'olfatto. Parte prima, Memoria prima. *Arch. Ital. Sci. Farmacol.*, 1935, **4**, 1-19.

102. ———. Contributo allo studio fisio-farmacologico dell'olfatto. Parte seconda, Memoria prima. *Arch. Ital. Sci. Farmacol.*, 1935, **4**, 77-115.

103. ———. Modificazioni della concentrazione solfocianica nel muco nasale ed in altre secrezioni, in relazione con stimolazioni del trigemino e dell'olfattorio. *Boll. Soc. Ital. Biol. Sper.*, 1935, **10**, 431-432.

104. ———. Relazioni fra stimoli semplicemente odorosi o odorosi-irritanti e concentrazione solfocianica del muco nasale. *Boll. Soc. Ital. Biol. Sper.*, 1935, **10**, 429-431.

105. ———. Contributo allo studio fisio-farmacologico dell'olfatto. Parte prima, Memoria seconda. *Arch. Ital. Sci. Farmacol.*, 1937, **6**, 241-278.

106. ———. Contributo allo studio fisio-farmacologico dell'olfatto. Parte prima. *Arch. Ital. Sci. Farmacol.*, 1940, **9**, 41-51.

107. ———. L'olfatto. Ipotesi che tendono ad interpretare il meccanismo di percezione degli odori, e ricerche destinate ad avvalorarle. *Arch. Psicol., Neurol., Psichiat. e Psicoterapia,* 1942, **3**, 31-78.

108. ———. Elementi olfatto-attivi di fronte al grado occupato nella scala delle electronegatività, e sostanze odorose di fronte al tipo ed alla intensità delle energie di legami. *Rivista Clin. Med.*, 1946, **46** (Suppl. I, *Marg. Otolaryng.*, v. 5, F. I) 10-16.

109. Pauling, L. The Nature of the Chemical Bond. (2d Ed.) Ithaca, N. Y.: Cornell Univ. Press, 1944. Pp. xvi+450.

110. ———. Analogies between antibodies and simpler chemical substances. *Chem. Engng. News,* 1946, **24**, 1064-1065.

111. ———. Molecular architecture and biological reactions. *Chem. Engng. News,* 1946, **24**, 1375-1377.

112. Pfaffman, C. Gustatory afferent impulses. *J. Cell. Comp. Physiol.*, 1941, **17**, 243-258.

113. ———. Taste and smell. Chap. 29 in Stevens, S. S. (*Ed.*) *Handbook of Experimental Psychology.* New York: Wiley, 1951. (Pp. 1143-1171.)

114. Prelog, V., Ruzicka, L., Meister, P., & Wieland, P. Untersuchungen über den Zusammenhang zwischen Konstitution und Geruch bei Steroiden. *Helv. Chim. Acta,* 1945, **28**, 619-627.

115. Prelog, V., Ruzicka, L., & Wieland, P. Über die Herstellung der beiden moschusartig riechenden Δ^{16}-Androstenole—(3) und verwandter Verbindungen. *Helv. Chim. Acta.*, 1944, **27**, 66-71.

116. Pressler, C. E. Theories on odors. *Drug and Cosmetic Ind.*, 1948, **62**, 180-182, 273-274.

117. Riccitelli, E. La theoría sintónica de la visión; nuevas orientaciones en la interpretación del mecanismo de la olfacción. *Rev. Soc. Argent. de Otorrhinolar.* (Buenos Aires), 1925, 1, 88-111.

118. Rideal, E. K. Introductory address, Part II: Physico-chemical aspects. In Faraday Society, *Modes of Drug Action.* London: Gurney & Jackson, 1943. (Pp. 368-372.)

119. Ruzicka, L. Die Grundlagen der Geruchschemie. *Chem. Ztg.*, 1920, **44**, 93-94, 129-131.

120. Saksena, B. D. The complete Raman spectrum of glycerine. *Proc. Indian Acad. Sci.*, 1939, **10A**, 333-340.

121. Sellari-Franceschini, A. Sulle possibili relazioni fra acutezza olfattoria e concentrazione solfocianica del muco nasale. *Boll. Mal. Orecchio.*, 1950, **68**, 599-615.

122. Sellari-Franceschini, A., & Camaiori, L. Le variazioni dell'acuità olfattoria e della concentrazione solfocianica nel muco nasale in gravidenza e puerperio. *Boll. Mal. Orecchio,* 1950, **68**, 488-507.

123. Sexton, W. A. Chemical Constitution and Biological Activity. London: Spon, 1949. Pp. xxi+412.

124. Singh, B. K., & Lal, A. B. Studies on the dependence of physiological action on chemical constitution: I. Difference in odour of d-, l-, and dl-derivatives of amino- and bisamino-methylenecamphors. *Proc. Indian Acad. Sci.*, Sec. A, 1940, **12**, 230-234.

125. v. SKRAMLIK, E. Über die zur minimalen Erregung des menschlichen Geruchs bzw. Geschmackssinnes notwendigen Molekülmengen. *Pfg. Arch. Ges. Physiol.*, 1948, **249**, 702-716.

126. STEFANINI, A. Sur la théorie de la perception des odeurs. *Arch. Ital. Biol.*, 1924, **74**, 8-17.

127. TAYLOR, N. W. A physico-chemical theory of sweet and bitter taste based on the properties of the plasma membrane. *Protoplasma*, 1928, **4**, 1-17.

128. TEUDT, H. Eine Erklärung der Geruchserscheinungen. *Biol. Zbl.*, 1913, **33**, 716-724.

129. ———. Erwiderung auf Hellers Artikel "Zur Geruchtheorie von Teudt." *Biol. Zbl.*, 1920, **40**, 259-267.

130. TROLAND, L. T. The Principles of Psychophysiology, Vol. 2. New York: Van Nostrand, 1930. (Pp. 261-280.)

131. TSCHIRCH, A. Die Riechstoffe und das Riechen. *Chem. Zbl.*, 1921, **3**, 190-191.

132. ———. Une théorie chimique de l'olfaction. *Chim. et Ind.*, 1923, **9**, 125-126.

133. TURK, A., & BOWNES, K. Inadequate stimulation of olfaction. *Science*, 1951, **114**, 234-236.

134. WATSON, E. R. Smell. *Biochem. J.*, 1922, **16**, 613-618.

135. WOKER, G. The relations between structure and smell in organic compounds. *J. Phys. Chem.*, 1906, **10**, 455-473.

136. WOOLLEY, D. W. A Study of Antimetabolites. New York: Wiley, 1952. Pp. xiii+269.

137. WOODWORTH,, R. S. Experimental Psychology. New York: Holt, 1938. (Pp. 477-500.)

138. YOUNG, C. W., PLETCHER, D. E., & WRIGHT, N. On olfaction and infra-red radiation theories. *Science*, 1948, **108**, 411-412.

Department of Psychology
University of California
Los Angeles 24, California

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 243-250.

ENGLISH AND AMERICANS: VALUES IN THE SOCIALIZATION PROCESS*

Department of Psychology, University of Connecticut

MAURICE L. FARBER

A. THE PROBLEM

It is by now almost a methodological tradition among students of culture and personality to focus upon the process of socialization of the child. While such a focus has on occasion neglected institutional factors acting upon the adult and limited itself doctrinairely to orthodox psychoanalytic factors, there can be little doubt that influences of massive importance have been revealed by this approach.

The socialization process can be directly studied by observation of the interaction of the child with his parents and associates, a method presenting formidable difficulties. Another possibility is to study the stated values and expectations of adults in a society with regard to childhood. To be sure, such verbalizations may well include projection, rationalization, and perhaps reaction-formation, but it is nevertheless a plausible assumption that they will also embody dominant and formative values of the society. It is this latter approach which is employed in the present study of English and American cultures.

The technique, similarly applied in the two countries, consisted in submitting to adults for completion, as a projective device, the following incomplete sentence:

"A properly brought-up child should be . . ."

All the subjects were insurance clerks, based upon the rationale that the selection of comparable occupational groups would control class and other subcultural differences and leave a relatively "pure" residue of national differences. The specific occupation of insurance clerks was selected because it has a similar position in the structures of the two nations, a situation that does not obtain for a number of other occupations such as the civil service which has a higher status in England than in America. Data were obtained from 32 cases in England and 81 in the U. S. The method

---

*Received in the Editorial Office on April 29, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

of the study as well as problems of the sampling are discussed in detail in a previous article (2).

B. BASIC DATA

The qualities given in the replies were categorized into those expected to be predominantly British and those expected to be predominantly American on the basis of hypotheses derived from the writer's previous study (2). The hypotheses are as follows:

1. The British in the socialization of their children aim at a non-threatening personality structure through the strong suppression of antisocial impulses by inner controls. The main anxiety is that the forbidden impulses will break through the controls.

2. The American socialization process, it is hypothesized, aims at a smoothly functioning individual, sensitive to interpersonal expectations, possessing a variety of skills enabling him to cope efficiently with his social environment. The main anxieties are that spontaneity and individuality might be crushed in the process and that values like sincerity and honesty might be compromised.

The actual findings, which support these hypotheses, are presented in Tables 1 and 2.

Although the number of mentions of any particular quality is generally too small for differences to be statistically significant, the totals of the qualities hypothesized for the two nations are different at a high level of statistical significance. Thus, of the total classified qualities mentioned by the British, 76 per cent fall into the group hypothesized as British, while only 37 per cent of the qualities mentioned by the Americans fall into this group. Of the total classified qualities mentioned by the Americans, 63 per cent fall

TABLE 1
PERCENTAGES OF BRITISH AND AMERICAN RESPONDENTS LISTING VARIOUS QUALITIES AS DESIRABLE IN A PROPERLY BROUGHT-UP CHILD

Qualities hypothesized as predominantly British Quality	British	American
Good manners, politeness, courtesy	53	43
Obedience	31	20
Thoughtfulness about others	16	12
Self-reliance	19	1
Create no nuisance	16	0
Hold his own without aggression	6	0
Self-control	6	0
Kindness	13	2
Kindness to animals	9	0
Total number of respondents	(32)	(81)

TABLE 2
PERCENTAGES OF BRITISH AND AMERICAN RESPONDENTS LISTING VARIOUS QUALITIES AS DESIRABLE IN A PROPERLY BROUGHT-UP CHILD

Qualities hypothesized as predominantly American		
Quality	British	American
Respect for parents, adults, authority	19	30
Sincere, honest	12	22
Getting along with others, mixing well	0	11
Unselfish	0	9
Well-behaved	3	7
All-aroundness	0	4
Individuality, originality	6	15
Neatness	3	6
Influenced by religion	6	11
Loving	3	6
Feels loved by parents	0	4
Not spoiled	0	4
Genial, good-natured	0	2
Intelligent	0	2
Taught about sex	0	2
Total number of respondents	(32)	(81)

into the group hypothesized as American, while 24 per cent of the qualities mentioned by the British are to be found in this group.

C. SOCIALIZATION OF THE BRITISH CHILD

The pattern which appears to emerge reveals a strong attempt to suppress impulses which are socially disturbing. Thus, for example, Subject B4 states "He should not plague the life out of his parents," Subject B23 would like him to be "unobtrusive," and Subject B12 that he "not be cheeky." Subject B8 states emphatically: "This is a must upon which I feel deeply: he must be well-behaved at meals." Very little feeling of this sort was expressed by the Americans. The category in the statistical table "Create No Nuisance," in which the British replies outnumber the American, exhibits this suppressive orientation directly while the categories "Good Manners, Politeness, Courtesy," "Thoughtfulness about Others," and "Kindness to Animals" reveal its more positive manifestations. "Kindness to Animals," which is never mentioned by the Americans, is of rather special interest, having been noted by informal observers as being peculiarly British. It is possible, if one may be allowed to speculate, that animals, as has been suggested by Rorschach experts, stand for childish impulses, and that the kindness represents a nostalgic sympathy for the impulses in oneself which must be suppressed by society.

"Obedience," which is mentioned with considerable frequency by the

Americans as well, would seem to involve both the desired end product of a non-threatening organism and the process for producing one. British Subject B7 states, for example, "Too many parents make the mistake of trying to reason with a very young child."

"Self-reliance" as stressed by the British respondents may be seen, similarly to the others, as a quality of individuals who do not constitute an overly-dependent burden upon their fellows. British Subject B9 spells it out: ". . . must be able to look after itself when traveling and away from home."

There is some evidence, though the number of replies is small, that the steering mechanism for this desired personality type is an inner one, i.e., "Self-control," through an internalized set of values. In psychoanalytic terms this would involve a strong Super Ego formation.

The major dilemma of socialization in England is to achieve a tight suppression of antisocial impulses which will somehow not break through the inner barriers. It is about this potential danger that anxiety exists. Such anxiety tinges the statement of Subject B25: ". . . to be able to hold his own without aggression."

D. Socialization of the American Child

The American pattern aims at a smoothly functioning individual, equipped for getting ahead with a varied armament of social skills. Thus, Americans mention more frequently such qualities as "Intelligence," "Neatness," "Geniality and Goodnaturedness," and "Knowledge of Sex."[1] At the same time, it is desired that the child be an "All-around" one, perhaps partly because such an individual is more genial but also because a flexible ability to shift skills is required in order to succeed. Respondent A3 feels that the child should have "a well-rounded schedule of activities, from Sunday School to sports and from helping his parents to playing amiably with both Johnny and Mary."

Yet these special skills are perhaps all subordinate to the broad all-important one of "Getting Along With Others," a quality not mentioned at all by the British. Moreover the demand is rather for *behavioral* adjustment than that any particular feelings be connected with it. "Being Well-behaved" is more frequently mentioned by the Americans. Even such a statement as that of American Respondent A14, that a child be "cheerfully obedient," seems less concerned with the feeling than the external appearance.

---

[1] The absence of any mention of sex by the British is congruent with their greater suppression of impulses.

An important difference between Americans and British emerges at this point. It is that the American stress is upon adjustment of the child *to other children,* while the British are much more concerned with their behavior toward *adults.* Thus, American Respondent A43 speaks of "being able to play together with other children in harmony" and American Respondent A14 of his being "well-adapted to his own age group," while no similar sentiments are expressed by the British. The consequences of this difference in attitude will be discussed later in this paper.

The impression emerges that the locus of control among the Americans is, in a way, more external than in the British. Thus, there is stress upon "Respect for Parents, Adults, and Authority" and being "Influenced by Religion."

An important American anxiety is that socialization will go too far and destroy individual impulses. For example, Respondent A43 says "Discipline is necessary but it must not stifle the child's growing individuality," Respondent A61 that "He should be well-behaved but not too angelic," Respondent A68 "respectfully and politely self-expressive," and Respondent A59 ". . . to some extent a rascal."

Anxiety is probably also involved in the American insistence that the child be "Sincere, Honest" and "Unselfish." The recognition that these qualities may not indeed always be the most efficient ones in achieving success and may at times be lacking in the behavior of oneself and one's associates seems to result in an over-compensating stress upon them, a situation not obtaining among the British.

Upward social mobility, which is implicit in the high valuation of social skills discussed above, is also more directly suggested, as in Respondent A2's statement that a child "should have an equal chance of succeeding." And yet a counter-pressure of moderation is involved, as in the statement of Respondent A7 ". . . adjust or content himself with what parents can offer, but later in years, try to gain more either in education or environment." *Adjustment,* i.e., adaptation of behavior to a situation, is indeed a peculiarly American concept.

Finally, *love* appears as an element in several American responses, both in the form of the child's "Receiving Love" and to his being "Loving." Our data on this point are too meager to indicate the possible dynamics. It might be noted in this connection, however, that Gorer (3) has maintained that for Americans love and success become equated, and that this develops an insatiable hunger for indications of love from others.

E. SOME CONGRUENCES

The findings of the present study, based as they are upon a relatively small number of cases, are better viewed as suggestive than as conclusive. If, however, it can be demonstrated that they possess certain congruences with material independently arrived at by other investigations then the degree of confidence with which one can accept them is enormously increased.

The strongly suppressive element of British socialization, with the accompanying anxiety that antisocial impulses may break through, is revealed also in Kaldegg's comparison of German and British boys (4), in which a projective technique demonstrated a marked disapproval by British children of "bullies," i.e., those in whom forbidden aggressive and domineering impulses have broken through to constitute a threat.

The contrast between British inner controls and the more externalized locus of control in Americans is to be noted as well in the work of Wolfenstein and Leites (8), who showed that the plots of British films more frequently involved inner conflicts of motives, while the American films involved conflicts with outside forces.

In a more recent paper Wolfenstein (9), on the basis of a comparative analysis of the *Infant Care* Bulletin of the United States Children's Bureau for 1914 and 1942-45, notes a change from an earlier dread of strong and dangerous impulses in the child to a current view of infant impulses as harmless, which need no suppression but only to be safely diffused into a variety of activities. There is some resemblance of the earlier American attitudes to the present British ones, suggesting that a divergence has taken place between the two cultures. With regard to present-day America, however, our own findings would suggest, not a random diffusion of impulse, but rather a sharp channelization into skills and sensitivities for control over the environment.

Of considerable interest is the relation of the present material to the work of Riesman (7), whose hypothesis, roughly stated, involves a shift in American society from the inner-directed man, whose behavior is governed by values internalized early in life to the other-directed man, who is steered by the values and expectations of his peer-group. The inner-directed man cannot strictly be identified with the British character aims as described in the present study, but there can be no doubt that elements of inner-direction are to be found in the somewhat more complex constellation here described. Similarly, Riesman's other-directed man is not inconsistent with some elements of the American character aim as described here: the stress upon getting along with others and mixing well, upon being "all-around" and upon "respect" for elders and authority.

Entirely consonant with our data are the impressions of Margaret Mead (6) that English girls, in their handling of men's sexual interest, depend heavily upon the self-control of the English men, while American girls usually expect physical advances but possess techniques for controlling them.

Bateson (1) has pointed to a difference between adult males in the two nations, most sharply evident, perhaps, in the behavior of lecturers. The Englishman talks quietly, from a position of assured strength, while the American is louder, more spectacularly bent upon making an impression. Bateson explains this difference on the basis of the rôles of the children in the two families: The British child is expected to sit quietly and listen respectfully to the quietly assured statements of the father, while the American child is himself encouraged to perform and is applauded for the brightness and charm of his performance. Leites (5) has noted that this analysis involves a basic difference in identification, i.e., the British child must at some point switch from the listening child rôle to that of the quietly assured father, whereas the American child does not switch but remains the child stridently trying to impress. The findings of the present study would appear to offer a possible explanation: the British child's behavior is strongly oriented toward not discommoding the tacitly superior adult world. The pressures are thus toward stepping into the preferable adult rôle in which he will become the central family figure, as soon as this is acceptable. The American child, in contrast, is more oriented toward his peer group of other children and in his contact with adults is rewarded for the "cute," spontaneous attention-getting childishness of his behavior. The adult state is not implicitly held up as the ultimately desirable goal. He already is the central family figure. Learning theory would indicate that with such reinforcement the tendency would be to retain the childhood rôle and behavior.

It will be recalled that the primary data upon which this paper is based are concerned with values and expectations concerning children. The socialization process in a society, however, has as its aim the production of certain kinds of adults. We may assume, therefore, a continuity, between these findings and the adult character structure in the two nations. We have indeed, if any be needed, evidence for this in the previous study by the present writer (2) in which the question concerned qualities admired most in adults and where the replies show considerable overlap with those considered desirable for children.

REFERENCES

1. BATESON, G. Morale and National Character. In Watson, G. (*Ed.*), *Civilian Morale.* New York: Reynal & Hitchcock, 1942.
2. FARBER, M. L. English and Americans: A study in national character. *J. of Psychol.,* 1951, **32**, 241-249.
3. GORER, G. The American People. New York: Norton, 1948.
4. KALDEGG, A. Responses of German and English secondary school boys to a projection test. *Brit. J. Psychol.,* 1948, **39**, 30-53.
5. LEITES, N. Psycho-cultural hypotheses about political acts. *World Politics,* 1948, **1**, 102-119.
6. MEAD, M. The application of anthropological techniques to cross-national communication. *Trans. New York Acad. Sci.,* 1947, **9**, 133-152.
7. RIESMAN, D. The Lonely Crowd: A Study of the Changing American Character. New Haven: Yale Univ. Press, 1950.
8. WOLFENSTEIN, M., & LEITES, N. Movies: A Psychological Study. Glencoe, Ill.: Free Press, 1950.
9. WOLFENSTEIN, M. The emergence of fun morality. *J. Soc. Issues,* 1951, **7**, 15-25.

Department of Psychology
University of Connecticut
Storrs, Connecticut

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 251-269.

PERSONALITY TRAITS OF PSYCHOTIC PATIENTS AS REVEALED IN THEIR SPONTANEOUS PAINTINGS*

Department of Psychology, Pennsylvania College for Women

EDGAR M. FOLTIN

A. INTRODUCTION

In 1943 while engaged in the supervision of students doing field work at the Eastern State Hospital in Williamsburg, Virginia, the author adventitiously came upon a large collection of paintings freely produced by female psychotic inmates.[1] In studying these productions he became interested in the question whether, quite apart from their content, the form and execution of such paintings could be made to yield a key to the personalities of their authors. Pertinent literature frequently contains assertions of the existence of general relationships between various characteristics of paintings and the personality traits of the artist. Thus it has been claimed that schizophrenics tend to intersperse their drawings with written characters, that they tend to prefer symbolism in their artistic creations, that they tend to work many minute details into their productions, and that, if not in an excited state, they tend to show a preference for somber colors. In the publications of Reitman (14, 15), Napoli (11), Zimmerman and Garfinkle (19), Pappenheim and Kris (12), and Martin and Weir (8), to mention a few examples, such claims occur in great number.

If such relationships can be demonstrated to exist, the presumption immediately comes to mind that they could be used for diagnostic purposes and might, in this connection, have considerable value. It seems, however, that before much confidence can be placed in these findings another important step needs to be taken in order to establish their validity. It would be necessary to know how close these relationships actually are, and to what extent they are exclusive. In other words, it would be necessary to

---

*Received in the Editorial Office on April 30, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to take this opportunity to express his appreciation to the authorities of the Eastern State Hospital, especially to Dr. Maxwell-Beamer, for their understanding coöperation in this study; to Dr. Alfred Dietze, my colleague, for his advice on questions of English style; and to Pam Pauly, Earl Aluisi, Elaine Dailey, Joe McCrary, and P. O'Keefe, who as graduate and undergraduate assistants ably helped in the supervision of the paintings and in the tabulations and computations upon which these findings are based.

establish that these relationships appear significantly more often than may be expected by chance and, in addition, that there is a significant difference in the painting behavior of patients manifesting certain syndromes or symptoms and those who are free from these conditions. To illustrate more concretely, if it appears that depressed patients show a preference for painting in black, before this can be accepted as a sign diagnostic of depression it would have to be shown that this is not merely a chance phenomenon, and that, furthermore, such a preference does not appear in similar degree in the paintings of non-depressed patients.

Only by statistical analysis can these problems be solved. Unfortunately, however, there seems to be an inverse relationship between the frequency of claims for the existence of the above type of correlations and attempts at their statistical corroboration. A paper by Robertson (17), presents an outstanding exception. Earlier Anastasi and Foley (1) made a clear-cut statistical comparison between the paintings of psychotics as a group and those of normal individuals, and report significant differences. Their findings, however, raise a further question, viz., which of the various psychiatric groups comprising the category "psychotic" contributed most to the differences they discovered?

It is to get more information on the latter problem that the present investigation was undertaken. As a first formulation, it seemed appropriate to analyze our data according to the separate psychiatric groupings into which our patient-painters had been classified in the hospital records, i.e., to treat schizophrenics, manic-depressives, etc., as separate groups. Further considerations, however, emphasized the diversity of the materials in such categories. For instance, the general behavior of a patient who goes from the manic phase of his disorder to the depressive phase changes remarkedly. It would, therefore, be not at all surprising to find that his painting behavior also undergoes a corresponding change. Consequently, if we were to work with manic depressive patients as an undifferentiated group and discover significant characteristics in their painting, we would still be faced with the question concerning the extent to which these characteristics are attributable to the manic and, respectively, depressed states in which the paintings were produced. Similar difficulties would be encountered in dealing with schizophrenics as an undifferentiated group. Thus, step by step, the original problem regarding the psychopathological personality in general was narrowed down to the possibility of discovering relationships to syndromes, and finally to more specific symptoms. In short, it was decided to investigate the question whether any statistically reliable relations exist between psychotic syndromes, or, especially and more specifically, symptoms and certain identifiable formal characteristics in the freely produced paintings of patients.

B. Description of the Sample

Before beginning the experiment, a study of the collection revealed the necessity for making some kind of selection in the paintings in order to avoid the biasing influences of certain factors. First it was decided to reject all paintings bearing no name or date, and also, in order to avoid the influence of the sex factor the few produced by male patients. Second it was decided to accept the work of only those patients who had painted at least four pictures on different days. This was done in an attempt to gain a more representative sample and avoid chance paintings, that might not be characteristic of the patient but be due to temporary fortuitous circumstances, whether occasioned by environmental or internal happenings. In the case of female patients in particular it seemed desirable to avoid the chance that the use of a single picture would have been painted during a critical period of the monthly cycle. The larger the sample of pictures chosen the greater the probability of obtaining a representative distribution over a longer period of time. An insistence on too large an individual sample would bring about too great a shrinkage in the total number of patients includable in our investigation so that statistical analysis would become of doubtful value, if not impossible. A minimum of four was believed by inspection of our collection to be the highest practicable figure to insure a satisfactory sample.

After applying these criteria of selection there remained 1,300 usable pictures painted by 74 female patients. These were distributed as follows: 45 belonged into the various categories of schizophrenia, 11 were manic-depressive, 4 suffered from involutional melancholia, and 14 belonged to various other categories, or were classified with doubtful diagnoses or undiagnosed. It should also be pointed out that these patients do not constitute a random sample from the whole population of females in the hospital. They had been selected by the physicians from the point of view of their possible suitability for occupational therapy. Patients not able to make their way across campus to the painting room, greatly disturbed, confused, destructive, stuporous, strongly aggressive or overly negativistic, fugitive and similar cases are not included in the sample.

In the practice of the hospital the majority of newcomers to occupational therapy was usually deliberately guided towards painting as a beginning. If they did not like the idea, they were persuaded to try. In such cases the painting activity was, of course, not entirely spontaneous. The content of the picture produced, however, was always spontaneous, since no suggestions were made by the nurses what the patients were to paint. Painters were never criticized or guided in any way. In respect to content and the execution, we may, therefore, regard our paintings as spontaneous productions.

The pictures selected for our research were executed in poster paint. Each painter was provided with 12 pigments from which she could make her choice, black, white, light and dark red, light and dark blue, light and dark green, yellow, brown, orange, and purple. Paper of fairly uniform size, 45 cm. by 60 cm. was provided. Every patient had a No. 2 sable brush, a rag to wipe it clean, and a can of water. The patients were always asked to clean their brushes before going from one color to the next in order to avoid as much as possible the inadvertent mixing of the colors supplied. Each patient received a fresh supply of paint at each session. If she used her whole supply of a given color, the cup would be refilled.

The number of pictures contributed by individual patients varied greatly between the required minimum of four and an extreme of 189, the median of the distribution being approximately eight. The four extreme cases fall into the dementia praecox group, two being diagnosed as hebephrenic and two as catatonic; however, division of painters into two groups above and below the median number painted results in an even distribution of schizophrenic and non-schizophrenic patients. It seems unjustified, therefore, to consider in our case the possibility of a special spell of high creativity in schizophrenic patients.

C. PROCEDURE

A chi-square design involving four-fold contingency tables was selected as an appropriate method of analysis. This involved as a first step the selection and definition of the categories to be investigated, viz., the characteristics of the paintings and the symptoms of the patients with which they might be correlated. The present section describes the procedure followed in the selection of categories as well as the statistical methods used.

1. *Selection of the Characteristics of the Paintings*

The first task was the selection of characteristics of the paintings to be investigated. An extensive study of the whole collection of pictures resulted in the selection of a list of 111 characteristics against which the 1,300 paintings were to be checked. The items were carefully defined and fell roughly into four categories as follows: (*a*) *Basic characteristics* entering into the composition of all pictures, such as pigment selected, brushwork, etc. (*b*) *Single items* repeatedly found to be parts of the total pictorial composition, such as objects, script, certain figures, etc. (*c*) *Style,* i.e., the characteristic mode in which the painter expresses himself in his painting taken as a whole. (*d*) *Design and general appearance,* i.e., attributes of the overall treatment of the picture, such as the amount of integration it shows, distribution of color, mass, and line, general tidiness, etc.

The selection of the different items in such a list, and their definition, is admittedly arbitrary. There is no recognized logic that compels the selection of one item and the rejection of another. Such arbitrariness does not, however, constitute a source of error since the items selected do not have any significance in and of themselves. It is only in the crucible of the statistical analysis to which they may be subjected, that any inherent significance or lack of significance may be revealed.

Tally sheets were prepared listing the characteristics above described and the painting traits of each patient were tabulated on a separate such *trait-sheet* bearing a code number for identification of the patient. The author and an assistant, working side by side, examined all the pictures of each patient as a group, and discussed them from the point of view of the judgments to be made with the objective of reaching agreement on each item in the list, i.e., agreement regarding the presence or absence in the picture of each trait in turn. No entry was made where an agreement could not be reached. After the first tabulation was completed, the procedure was repeated independently after an interval of four weeks. Only eight discrepancies occurred between the first and second rating and these were rejected in the final tabulations.

The number of items entered on the sheets of different patients varied considerably. Some patients had as few as 15 entries; others as many as 51. These individual differences are easily explainable. In many cases an entry was not made because the painter did not show the trait in question. In other cases no entry was made because it was a matter of doubt whether the trait was present or not; for instance, in the case where a painting was not distinct enough, or too poor in realism to permit a clear-cut decision as to whether or not a certain item was present.

2. *Selection of Symptoms*

A second task involved the determination of the specific symptoms to be studied. Some of the patients in our population had died or had left the hospital since their painting activity; others had undergone considerable change in their disorder in the direction of improvement or becoming worse. Since the purpose of our investigation was to compare the personality of the painter at the time of his work with his painting, it became necessary to depend entirely on the case history records containing the anamneses, mental examination, staff conference entries, and general progress notes as a basis for such reconstruction. Since the data on the patients were intended for statistical analysis, it became necessary to select symptoms and counter indications that were mentioned in the case histories with sufficient frequency and with a reasonably precise and consistent terminology. With this in mind, large numbers of case histories were examined at random, netting 59 usable symptoms and counter indications.

Individual *symptom-sheets* were now prepared for our 74 patients. Each case history was carefully searched twice for the symptoms listed. Symptoms were tallied whenever the case history used the exact term or phrase that we had selected, or whenever we found a description of the patient's behavior that left no doubt of the identification of the symptom. Only medical observations made within the hospital were tabulated. Disregarded were, therefore, all symptoms described by the informant. Dr. Maxwell Beamer, one of the psychiatrists at the Eastern State Hospital, gave assurance that whenever a symptom was not mentioned in a case history, it could be taken for granted that it was not exhibited by the patient. In such cases a negative entry was made on the symptom-sheet.

Doubtful cases were not entered at all. If, for instance, the behavior of a patient was at one time described as friendly, and at another time as hostile, neither observation was used, since there was no way to determine in such cases the probable states of a patient at the time of her paintings. Entries made one or more years prior to the painting were only accepted if later notes reported condition unchanged, or if the psychiatrist in charge of the case could make a statement to this effect when queried. Finally, entries in case histories were rejected whenever they were qualified by words, such as probably, slightly, somewhat, occasionally, and the like. This selective process led to a considerable, though unavoidable, loss of material. There were case histories that yielded *symptom-sheets* with as few as 36 entries, the best one contained 53.

3. *Setting Up Contingency Tables and Calculations of Chi-Squares*

Before the contingency tables necessary for chi-square analysis could be expeditiously set up, two preliminary steps seemed necessary to handle this mass of material without IBM equipment. First the information entered in the *trait-sheets* and *symptom-sheets* was transcribed to a set of *cross-sheets* set up for each of the 59 symptoms in pairs, one of each pair for the presence, the other for the absence of the trait in question. The row captions indicated the painting traits, while the column headings indicated by code number the patients for which entries were made. The cells were marked by check marks to indicate the traits possessed by each patient and the number of tallies in each row was totalled in the right margin.

The second preliminary step involved a transcription of these data to a master-sheet by again listing the painting traits in the row captions and using the column headings to indicate the symptoms in double columns, one for the presence and the other for the absence of the trait shown. This provided 59 x 111, or 6,549 double entries in the cells indicating for each positive or negative occurrence of a symptom the frequency of each associated painting trait. Thus, from the *master-sheet,* it was, for instance, possible, to read directly how many demented and how many non-demented patients used *inappropriate* colors, and by subtractions from the total number, how many did not show this trait. Each of the double cells, therefore, yielded the necessary data for setting up a four-fold contingency table for calculating chi-squares; and the total number of such cells gave the maximum of comparisons that could be made on the basis of our set of categories, i.e., 6,549.

This method of setting up contingency-tables for analysis, it should be remarked, uses patients who were free of the symptom under investigation as a control group. Thus, for instance, fearful patients were pitted against the non-fearful ones. The question might be raised whether it would not have been more advantageous to use non-psychotic subjects as controls. It would easily have been possible to obtain a collection of paintings by normal subjects and to analyze them according to our list of painting traits. Such an outside control group would have had the statistical advantage of increasing the grand total of cases in each four-fold table. Nevertheless, it was deemed preferable to use an inside control group for two reasons. One is the tacit assumption in using an outside control group that all who compose it are free from the symptoms that were selected for investigation. Yet, when it comes to symptoms like coöperative, social, remote-memory-bad, etc., there is reason to doubt such an assumption. In any case it would seem that a person under psychiatric observation who is declared free from a symptom is much more likely not to have it than a "normal" person for whom we have not even the results of a perfunctory mental examination.

The main reason, however, for preferring an inside control stems from another consideration: for in using an outside control group a variable is introduced that might have profound influence on the spontaneity of painting, viz., the factor of institutionalization. Two groups might be well matched in regard to sex, race, age, education, and even social background, but there would still remain this important uncontrolled factor to becloud the issue. It is well known that enforced institutionalization, as in penal institutions, sometimes affects the personality to such an extent that a neurotic condition or even a psychosis may develop. Even when institutionalization does not have such severe consequences, its possible impact on the personality cannot be ignored. In comparing the spontaneous paintings of institutionalized psychotics and non-institutionalized normals, it would be difficult to say that any of the differences that might appear are unequivocally linked to the psychotic condition and are not effects of institutionalization.

In calculating the chi-squares from the four-fold tables, the correction for continuity was always used, regardless of the number of cases entered into any one of the four cells, a procedure that is recommended with good reason by Edwards (2). For this reason a chi-square formula that already included this correction was used throughout (2). The 5 per cent level of significance was adopted for rejection of the null-hypothesis. Since work with a four-fold table allows only one degree of freedom, no chi-squares were accepted that turned out to be smaller than 3.84. Also rejected were all otherwise significant chi-squares in which any one of the expected frequencies was found to be less than 5.00. This is in accordance with the suggestion given, among others, by Edwards (3). In calculating the expected frequencies the procedure outlined by Johnson (5) was followed. Altogether 29 chi-squares, otherwise significant, were rejected on these grounds. Further reference to them, however, will be made later on. After all this sifting 15 chi-squares remained that yielded results significant at the 5 per cent level or better. These are listed and discussed in the section on results.

D. Control Experiment

The reliability of the 15 significant chi-squares that our original experiment yielded depends on three factors, namely the pictures painted by the patients, the symptoms demonstrated by the painter as recorded in the case histories, and, finally, the statistical procedure. Each one of these three items is open to criticism.

The paintings were by-products of occupational therapy and were, therefore, not produced under strictly controlled conditions. Although paper and brushes were uniform, there are some doubts as to whether the painter had at all times the same choice of color. It is possible, for instance, that a patient was out of one of the colors and continued to paint before the paint container was refilled. The painting room, too, was gaily decorated with pictures by different patients. Again since more than one patient sometimes painted at the same table it is possible that one might have copied ideas or style from her neighbor. Free from such uncontrolled external influences, the patient might have painted in a different way.

The entries in the case histories were not always what an experimenter would wish them to be. The Eastern State Hospital of Virginia, during the war and immediately afterwards, was, at times, badly understaffed and overcrowded. The progress notes in the case histories were, therefore, in some cases extremely scanty. They did not always fall even near the date of painting. Possibly the patient, at the time of the painting, did not labor under the mental difficulties which she demonstrated during the nearest mental examination. For our purposes only symptoms that were reasonably consistent throughout a case history were tabulated, but even so, disturbing doubts remain.

The transition from peace to war, and back to peace produced in its wake a rapid turnover in the medical staff with considerable variation to psychiatric training and experience of this personnel. This could have led to inconsistency or fluctuation in the quality of the psychiatric observations and in the terminology used in the case histories. Besides, it was found in another connection, quite apart from the hospital here involved, that the standards of different psychiatrists vary to a considerable extent (6).

Finally, our statistical calculations were based on small samples. Although a correction for continuity was employed throughout, and no chi-square was accepted that was based on an expected frequency of less than 5, some dangers are ever present when chi-square is used with small samples (9). Furthermore in evaluating our 15 significant chi-squares, it must be borne in mind that they are the meager harvest of 6,549 chi-square possibilities. The question arises whether our chi-squares might not be the product of mere sampling error.

These considerations do not justify drawing sweeping conclusions from our original experiment without the check of a control experiment. A new series of paintings was, therefore, collected, under the author's supervision and with as much control of the painting conditions as possible. Dr. Maxwell-Beamer selected a new group of female patients. Upon the author's request she tried to choose functional patients with a well established diagnosis and a not too scanty case history. Patients were avoided who had recently undergone shock therapy, and such treatment was never administered during the painting period. A quiet bare room was selected to which the nurses would bring the patients. There were never more than two patients painting at the same time and they were seated so that they could not observe each other's activity. Two assistants, both girls, were trained to direct the actual painting. When the patients arrived, the accompanying nurse withdrew. The painting materials had already been laid out; it was the same as in the original experiment. The patients were simply invited to paint whatever they wanted. No suggestion regarding the possible content of the picture was ever made. If a patient refused to paint, and could not be persuaded to coöperate, no pressure was used. Each patient had exactly one hour at her disposal during which she could paint as many pictures as she wanted. She was not allowed to continue on a picture beyond this time even though she had not finished her very first one. Every patient had to come to at least four different painting sessions spaced at least one week apart. Only patients were accepted for this control experiment of whom at least four completed or four started pictures were obtained, each painted during a different week.

A number of the 42 patients with whom this experiment was started, was lost. Some did not want to paint any more as the sessions progressed; others became severely disturbed or bed-ridden. A few were furloughed or paroled before they had gone through the minimum painting schedule. At the end there remained for use pictures done by 20 different patients.

It would now have been possible to analyze these new data by the same procedure that was employed in the original experiment. This would have yielded 15 new chi-squares for comparison with the 15 original significant results. This procedure, however, was not chosen for two reasons. First, with a total of only 20 patients the cell entries in a four-fold table would necessarily be too small to justify the calculation of chi-square. The second consideration, however, would have been decisive even if there had been a larger number of patients at our disposal. If general relations between certain symptoms and painting-characteristics exist, they have their practical value in the field of diagnosis. The mere analysis of a picture should tell us something about the patient. Instead of piling chi-squares on chi-squares, it seemed preferable to make the practically decisive test of prediction using the significant relations found in the original experiment as basis and the data of the control experiment as criteria.

With this in mind the new paintings were subjected to an analysis identical with that which had been used in the original experiment. The analysis was performed by the author, and by a graduate assistant familiar with this work. Neither of the judges had seen any of the patients during their painting periods, nor did they know their names. Where an agreement could not be reached between the judges in regard to a certain painting-characteristic, no entry was made on the *trait-sheet.*

Next the judges filled out for each patient a *prediction-sheet* showing the painting traits and the symptoms, respectively, in the rows and columns. Whenever a patient showed a painting trait, that in the original experiment showed a significant relation to a certain symptom, a positive entry was made in the appropriate cell. When she did not show the painting trait in question, a negative entry was made. These entries were in the nature of positive and negative predictions of which there was a total of 164. This procedure was now also used for the 29 painting-trait-symptom relations that yielded otherwise significant chi-squares but had been rejected because expected frequencies were found in one of the cells to be below five.

Still "blind" in regard to the identity of the 20 new painters, the judges now went through their case histories, using the same method as in the original experiment. Since all patients were still hospitalized, it was also possible this time to check older entries in the case histories by personal observation and interviews with the patients. When the experimenter's findings did not agree with the entries in the case histories, the point was cleared with the psychiatrist in charge of the patient. The symptoms present or absent were then entered on a *symptom-sheet* which was identical with that used in the original experiment.

Finally the identity of each patient was established and the predictions compared with the symptoms found. The result of these comparisons was tabulated on a *result-sheet,* separately for the correct and the incorrect positive and negative predictions regarding each one of the 15 painting-trait-symptom relations. The same procedure was also followed for the 29 rejected chi-squares above mentioned.

E. RESULTS

The original experiment yielded both negative and positive results.

1. Our procedure put into the statistical test tube at least a few of the claims that are found in literature. It so happens that our results do not support a single one of those we were able to check. There is first of all the time-honored statement that psychotic patients show a tendency to cover every bit of their paper or canvas with color and design (10). Out of 138 patients only 12 covered their paper completely with color and satisfied, thereby, at least one of the conditions for a *horror vacui.* This finding is all the more impressive because pigment covering in itself is not enough to

show this so-called fear of emptiness. We find also in literature the restrictive statement that it is not the psychotics in general but the schizophrenics among them who show this horror (12, 16). Comparing these patients with the rest of our sample, no justification is found for rejecting the null-hypothesis. The chi-square was 2.82, and with one degree of freedom, falls far short of the five per cent level of significance adopted as a criterion.

Other claims tested fare even worse. The old belief that the depressed use black more frequently than the non-depressed patients (7) finds not the slightest support. On the contrary, the non-depressed group showed, if anything, a greater proportion of black-users, although the difference was far from significant, the chi-square for this relationship being only 1.36. The claims that schizophrenics use inappropriate color (13), stiffness (12, 15), and repetitiousness of design (10), seem plausible enough if one thinks of their remoteness from reality, the stiff behavior of catatonics, and the mannerism so frequently exhibited by the hebephrenic patients. Yet, at least in regard to schizophrenics as a group, our findings yield no statistically significant relations. In comparing the schizophrenics with other patients, our chi-squares are in the order of 0.51, 0.45, and 0.18, respectively, far too low even to be considered. It has also been said that schizophrenics tend to disconnect what is linked together in reality (16). In comparing our schizophrenics with other female patients in regard to this painting trait, an insignificant chi-square of 0.98 is obtained.

2. On the positive side the original experiment yielded the 15 significant relations shown in Table 1.

Before drawing any conclusions from these findings a word should be said about the results of the control experiment. As stated earlier, this yielded 164 predictions utilizing the 15 significant trait-symptom relations. Of these 83 were positive and 81 were negative. It was known from the beginning that the positive predictions could not yield results that would indicate the value of the findings in our original experiment. We may note, for instance, that a painter lays his paint on the paper in a thin layer. We know, furthermore, from our statistical results that the fearful have a significant tendency to act in this way. A positive prediction that our particular patient, therefore, is fearful, may be in error, because patients who are excited also seem to prefer, to a certain extent, a thin use of paint, and the same holds true for confused patients who do not like to spread their paint thickly. Besides, it is conceivable that other symptoms that were not here investigated may also be correlated with the use of thinly spread pigment. What we can say, therefore, with a certain degree of confidence is

TABLE 1

SIGNIFICANT PAINTING—PERSONALITY RELATIONS

No.	Painting-personality relation	Chi-square
1.	Confused patients paint in an *untidy* manner, meaning they show a tendency to smear color over the part of the paper that was not covered by the design, make color blots, or execute the picture carelessly, as shown by painting over the outer contours of objects.	7.02
2.	Confused patients paint *without contrast,* meaning that their pictures as a whole show no tendency to set contrasting colors side by side without intermediary tones.	4.62
3.	Confused patients avoid *red-green contrast,* meaning that the colors red and green are not set side by side to any appreciable degree.	3.91
4.	Confused patients paint *white,* meaning that the slightest amount of the pigment white appears on at least one painting.	3.91
5.	Confused patients paint in a *weak* manner, meaning that pigment is predominantly used with a liberal admixture of water.	4.96
6.	Fearful patients paint *not off center,* meaning that the subject painted is not squeezed into one corner, or towards the extreme end of the paper.	4.66
7.	Fearful patients use *thin* color, meaning that pigment is spread on paper predominantly thinly, either because it is weak and only one or a few layers are used, or the color is strong and very little color is on the brush, or the brush is moved quickly and goes over the paper only once.	4.36
8.	Patients who show inappropriate reactions use, *not more paint,* meaning that more than half of the paper is left uncovered by pigment.	5.63
9.	Patients who show inappropriate reactions, use *single colors,* meaning that at least one picture is painted in one color only.	4.36
10.	Demented patients paint one or more *inappropriate objects,* meaning objects found in the pictures are out of place although they are not inappropriate in themselves. Examples: steps lead to a part of a house which does not show a door, or a table stands in the middle of a lake.	6.55
11.	Demented patients use at least once *inappropriate colors,* meaning that unrealistic colors are used in painting an object, as blue on faces.	4.26
12.	Seclusive patients use predominantly only *two or three colors* in painting their pictures.	8.49
13.	Seclusive patients paint *not multicolored,* meaning they do not predominantly use four or more colors in their paintings.	5.37
14.	Patients with a sloppy appearance paint a *base,* meaning, that along the bottom of a picture, or higher up, a horizontal colored band or line is painted with a more or less level upper rim. The objects in the picture stand on this bar as if on a support, but do not reach into it.	4.79
15.	Worried patients use *cross-bars,* meaning that at least once a cross-bar is drawn over doors or windows.	4.14

only in the nature of a negative prediction. Our conclusion can only be that a painter who does not use thin paint will probably not be fearful. The negative and not the positive predictions that we can make obviously form the test that will show whether our 15 chi-squares point not only seemingly but factually to significant relations.

TABLE 2

PAINTING-PERSONALITY RELATIONS REJECTED BECAUSE OF TOO LOW EXPECTED FREQUENCIES

No.	Painting-personality relation	Chi-square
1.	Confused patients do not paint in a *thick* manner, meaning that the pigment predominantly is not spread thickly on the paper either in one layer of rather undiluted paint, or in many weaker layers.	8.76
2.	Evasive patients use *overpaint,* meaning that they cover one painted object later with pigment in order to make it disappear from the picture.	8.22
3.	Excited patients paint with *thin* color (see definition in No. 7 of Table 1).	5.50
4.	Fearful patients use *weak* pigment (see definition in No. 5 of Table 1).	5.18
5.	Amented patients paint in a *muddy* way, meaning that they mix into one pigment white, brown, or black, and produce thereby a dirty effect.	4.73
6.	Demented patients do not show *complete use of paint,* meaning that they do not cover completely with paint, at least one of their pictures, not counting a little spot of free paper in one place or another.	4.61
7.	Patients showing a lack of initiative paint in a *weak* manner (see definition in No. 5 of Table 1).	8.73
8.	Negativistic patients paint in a *weak* manner (see definition in No. 5 of Table 1).	4.71
9.	Manneristic patients paint in an *untidy* manner (see definition in No. 1 of Table 1).	4.34
10.	Negativistic patients intersperse their paintings with *writing.* This trait was tabulated also if only single letters were found. Not counted as *writing* were only words added to a picture to name the painted object such as "house" at the bottom of a house, or the added name of the painter.	10.08
11.	Negativistic patients paint *circles and loops,* meaning that strokes of this kind appear in great number in their pictures so as to demonstrate a pronounced tendency in this direction.	5.55
12.	Negativistic patients paint *single-colored,* meaning they paint at least one picture in one pigment only.	4.61
13.	Negativistic patients paint *few details,* meaning that they are satisfied with abstractions, e.g., trees are painted as stems with a solid green hood for the leaves, or a face shows only the bare essentials of mouth, nose, and eyes.	4.26

In the light of these considerations it is not surprising to find that only 42 per cent of our positive predictions turn out to be correct. On the other hand in the case of 7 out of our 15 significant items the negative predictions were 100 per cent correct. These are: confused, *no red-green contrast;* confused, *white;* confused, *weak;* fearful, *not off center;* fearful, *thin;* inappropriate, *not more paint;* and seclusive, *not multicolored.*

Again, of all the 81 negative predictions made in regard to the 15 significant items, 66 or 81.48 per cent were correct. Testing the significance of this finding against the assumption of a chance distribution a chi-square of 32.10 is obtained. The corresponding value of *P* enables us to reject the null-hypothesis at better than the 0.1 per cent level of confidence.

In similar fashion we may examine the 29 significant chi-squares that we rejected because at least one of the expected frequencies was smaller than 5. An inspection of the negative predictions made regarding these 29 painting-trait-symptom relations reveals considerable differences in the accuracy of prediction from item to item. Thirteen comparisons, however, yielded 37 negative predictions that were 100 per cent correct. They might be worthy of further experimentation. They are listed in Table 2.

F. Discussion and Conclusions

Both negative and positive conclusions may be reached from these results.

1. Let us examine first those on the negative side. It was shown that not one of the sample painting-trait-symptom or syndrome relations claimed in literature and checked by the author found statistically significant support. It is true, of course, that other investigators speak in these cases of schizophrenic patients as such, while we deal only with female patients. It is hardly to be expected however, that an appropriate addition of male patients would have yielded acceptable levels of significance when the statistical picture for a female population is so far off the mark. The sample that we checked, although a small proportion of the many claims that have been made, was taken at random from the literature and failed to show a single significant comparison. In the light of this, it would seem justified to make the general statement that claims of painting-personality relations that are not based on statistical tests are apt to be highly unreliable. The reason why such statistically untenable claims are so frequently made, may, perhaps, be found in a quasi-halo-effect carried over from a few preliminary observations into later research, and in the tendency to overlook, or underestimate, the negative cases.

The claim for the use of *inappropriate colors,* however, deserves special consideration. It seems here that only the emphasis is misplaced. It is not schizophrenic patients as such who make the wrong use of color, but only that proportion of schizophrenics that is demented, as was brought out by our results. The demented patient, indeed, shows a tendency towards the use of *inappropriate color* that is significant at the 5 per cent level.

2. Let us now turn to our positive findings and briefly discuss a few of the statistically significant results that were obtained. It should be emphasized that these do not reveal relations existing between any individual

and his paintings, but rather group relationships, that is, correlations that hold good for our group of patients as a whole. Some of these relationships are easily understandable. Demented patients suffering from schizophrenia, e.g., show a tendency towards unusual responses in an association test, they are likely to put unrelated items together, and they perform poorly on a test involving the discovery of absurdities. Associative and conceptual thinking are very frequently impaired. This kind of mental disorganization is characteristic of the functional type of dementia (6). Taking these disorganized into consideration, it seems understandable enough that demented patients should also have a significant tendency to use *inappropriate colors,* i.e., to associate the wrong color with a given object. They put together in their paintings unrelated things without noticing the absurdity, thus acting in their paintings the same as they do in other situations that require associative or conceptual thinking.

A similar consideration makes understandable the general marked tendency for demented patients to use inappropriate objects. A demented patient who makes *word-salad* out of his linguistic expression may also be expected to put a motorcar into the middle of the ocean, since a similar dislocation of thought appears in both types of behavior. Because we universally find disjointed thinking in demented patients, the appearance of *inappropriate objects* in their painting becomes understandable. *Word-salad* and *pictorial-salad* seem both to point to the same basic mental disorganization.

Among the rejected items that yielded 100 per cent accurate negative predictions are included some that show an obvious link between symptom and painting trait, thus again making the generality of the relationship understandable. Negativism is the tendency to behave in a way opposite to a suggestion made. The negativistic patient would, in all likelihood, exhibit this inclination in the painting situation just as in any behavior. If asked to produce a picture, he would, perhaps, start writing instead of painting. When offered a choice of 12 colors, he would stick to a single one to show his opposition.

Venturing a little further afield to speculate about the probable rôle of symbolism, one may find at least two explanations for the painting of bars across windows by worried patients. It is possible that bars serve as a symbol of protection against the vague threats that the world constantly presents. It is, as in much symbolism, possible also to reverse the argument and hypothesize that bars symbolize the barrier of worries through which the patient feels unable to break to achieve his desired goals. No claim is made here for the truth of either explanation. They are simply mentioned as plausible hypothesis, although it may be stated that Mosse (10) expresses similar ideas with regard to the symbolic rôle of fences. He made the observation that fences symbolize either a protection or a limiting element standing between the patient and his goal.

We present these arguments to show that the generality and even universality of symptom-painting relations is not a bizarre and untenable concept. In the 15 plus 29 items showing a statistical relationship between symptom and painting, there remains, however, a number not open to an easy rational explanation. A few examples may appropriately be given for illustration. Thus it seems understandable that a confused patient should put white pigment on white paper, but at least to the present author, it seems difficult to understand why particularly the confused should not want to set green and red side by side in contrast. It seems hard to understand why particularly seclusive patients specialize in the use of a two or three color palette. It is equally understandable why patients with sloppy appearance should prefer to rest their objects on a baseboard or base line.

These puzzling quirks, however, do not seem quite so surprising if we take into account that a great proportion of patients in our sample belongs to the schizophrenic group and that schizophrenia is still the "mystery of psychiatry," as Strecker and Ebaugh (18) put it. In the case of other psychiatric disorders, presumably, we have a better chance to achieve a degree of insight, but even here we are far from complete understanding. In view of the poverty of our knowledge in the area of psychopathology it should occasion no great surprise to find that we do not have a ready explanation for all the personality-painting relationships that our analysis revealed. If, however, the existence of certain group relationships and the possibility of their generality is supported by means of examples that are open to a reasonable explanation, there is no good reason why we should reject statistically supported findings simply because there is no theory ready at hand that would account for and give meaning to such relationships.

In formulating our final conclusions it is necessary to avoid over-generalization. All that can be said is that group relations have been found to exist between specific painting traits and certain symptoms in a selected group of hospital patients. There is no good reason to believe that this phenomenon is restricted to the group that constituted our sample. Indeed, it is highly probable that the same phenomenon of relationships would be found if we were to study the painting behavior of male psychotics, of psychotics of different racial background, etc. In this sense it may be said that such relationships exist generally as a consistent correlate of specified symptoms, in other words that patients will reveal aspects of their personalities in their paintings as well as in other behavior.

It is a different matter however, to generalize about the specific relationships that we found. Our population consisted only of white adolescent and adult female psychotics in the Eastern State Hospital of Virginia. Similar relations may be expected in similar segments of the population of other mental hospitals in this country. This is all the more likely inasmuch as the Virginia hospital does not care exclusively for citizens of Virginia. This, however, is as far as analogy may be permitted to carry us. Other groups of psychotics possessing the symptoms selected for our study might conceivably exhibit different relationships. That would apply particularly to male groups, since Goodenough (4) has already reported differences in the paintings by boys and girls. It would be equally dangerous to extend our specific findings to a group of psychotic children, or to the category of non-white patients. Further research would be necessary before such conclusions could be drawn.

In other ways too, this investigation points towards the need for more research. Although tabulations of many symptoms and painting traits were made, all possibilities have not in any sense been exhausted. Other counts might yield new significant relations. It is equally likely that old counts that yielded probabilities near significance would prove significant, if repeated with larger numbers, and that rejected chi-squares would become acceptable. This holds true, of course, also for the claims of relationships that others have made in literature and which we were forced to reject as not significant, although it seems to the author that only the one concerning the *horror vacui* or complete paint covering, seems near enough to significance to warrant such hope. The field of personality-painting relations is still open to further statistical and experimental investigation and may yet yield results that will prove widely usable and helpful in diagnosis.

REFERENCES

1. ANASTASI, A., & FOLEY, J. P., JR. An experimental study of the drawing behavior of adult psychotics in comparison with that of a normal control group. *J. Exp. Psychol.*, 1944, **34**, 169-194.
2. EDWARDS, A. L. Experimental Design in Psychological Research. New York: Rinehart, 1950.
3. ———. On "The use and misuse of the chi-square test"—the case of the 2 x 2 contingency table. *Psychol. Bull.*, 1950, **47**, 341-346.
4. GOODENOUGH, F. L. Measurement of Intelligence by Drawings. Yonkers-on-Hudson, N. Y., World Book, 1926.
5. JOHNSON, P. O. Statistical Methods in Research. New York: Prentice-Hall, Inc., 1949.

6. LANDIS, C., & BOLLES, M. M. Textbook of Abnormal Psychology. New York: Macmillan, 1950.
7. LISS, E. The graphic art. *Amer. J. Orthopsychiat.*, 1938, **8**, 95-99.
8. MARTIN, A. W., & WEIR, A. J. A comparative study of the drawings made by various clinical groups. *J. Men. Sci.*, 1951, **97**, 532-544.
9. MCNEMAR, Q. Psychological Statistics. New York: Wiley, 1949.
10. MOSSE, E. P. Painting-analysis in the treatment of neuroses. *Psychoanal. Rev.*, 1904, **27**, 65-81.
11. NAPOLI, P. J. Finger-painting and personality diagnoses. *Genet. Psychol. Monog.*, 1946, **34**, 130-229.
12. PAPPENHEIM, E., & KRIS, E. The function of drawings and the meaning of the "creative spell" in a schizophrenic artist. *Psychoanal. Quart.*, 1946, **15**, 6-31.
13. PFISTER, H. O. Farbe und Bewegung in der Zeichnung Geisteskranker. *Schweiz. Arch. Neurol. Psychiat.*, 1934, **34**, 325-365.
14. REITMAN, F. Psychotic Art. London: Routledge & Kegan, 1950.
15. ———. Facial expression in schizophrenic drawings. *J. Ment. Sci.*, 1939, **85**, 264-272.
16. ———. The "creative spell" of schizophrenics after leucotomy. *J. Ment. Sci.*, 1947, **93**, 55-61.
17. ROBERTSON, J. P. S. The use of colour in the paintings of psychotics. *J. Ment. Sci.*, 1952, **98**, 174-184.
18. STRECKER, E. A., & EBAUGH, F. G. Practical Clinical Psychiatry. (5th ed.) Philadelphia: Blakiston, 1940.
19. ZIMMERMAN, J., & GARFINKLE, L. Preliminary study of the art productions of the adult psychotic. *Psychiat. Quart.*, 1942, **16**, 313-318.

Department of Psychology
Pennsylvania College for Women
Woodland Road
Pittsburgh 32, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 271-287.

SENSORY PATTERNING VERSUS CENTRAL ORGANIZATION*

Yerkes Laboratories of Primate Biology, Inc.

HENRY W. NISSEN[1]

A. THE PROBLEM

Many of the basic differences among explanatory accounts of behavior reduce to varying emphasis within each of two spheres of choice: (*a*) Either innate mechanisms or experiential factors may be assigned the major rôle in behavior determination. (*b*) Peripheral factors of stimulation and response may be given dominance in explaining behavior, or there may be primary emphasis on central organizing processes. Selections within these two areas of choice are often correlated. Those who stress the rôle of innate factors tend also to emphasize the importance of central organization. On the other hand, those who consider experience more important than native endowment usually put their main reliance on peripheral rather than on central determination. Historically, the first of these trends was represented by "faculty psychology" and by early instinct psychology. The other direction was taken by the theory of tropisms and by Watsonian behaviorism. In more recent times, the European ethologists and the Hullian S-R psychology exemplify these divergent views.

Stated so broadly, the issue is not very meaningful for psychological theory. It must be broken down into sectors, so that the implications of each view may be followed in detail and subjected to experimental test. One of the areas in which the influence of preference for peripheral or for central determination has been most apparent is in the interpretation of discrimination learning by animals. Within this area, there are at least three subissues: (*a*) the so-called continuity-discontinuity controversy; (*b*) the question of whether the concept of sensory patterning can satisfactorily account for certain complex discrimination behavior, or whether the postulated mechanism of the conditional or *if then* response provides a more adequate and economi-

---

*Accepted for publication by Karl S. Lashley of the Editorial Board, received in the Editorial Office on May 2, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] I am grateful to Professors K. S. Lashley and K. W. Spence for their valuable comments and criticisms of this paper. This of course does not imply that either of these critics is in complete agreement with the present discussion.

cal explanation; (*c*) the question of whether the facts of transposition are to be explained by the hypothesized mechanism of stimulus generalization; or whether, instead, animals discriminate on the basis of relational properties. In each case, the peripheralist holds to the first of the alternatives as presented here, whereas the centralist prefers the second one. In the present discussion we shall not attempt to deal with the first of these subissues, but shall limit ourselves to the second and third questions.

B. The Mechanism of Sensory Patterning

The peripheral view originally held that the response was associated with all stimuli impinging on the sensorium at the time. This assumption would account for the fact that several elements of a stimulus aggregate may each elicit the response. Very often, however, eliminating or changing even a small part of the total stimulating situation results in a diminution or absence of response. The concept of afferent interaction (8) was introduced in part to account for this fact. Several stimuli concurrently impinging on the sensorium influence each other, centrally, and it is only the products of the interaction which become associated with the response. Even a very simple discrimination habit may manifest afferent interaction. Consistent choice of one of two sizes, for instance, may be destroyed by changing the color, or some other contextual feature, in which those two sizes are presented. We may then say that the discrimination is one of stimulus-patterns: size-*A*-plus-color-*X* versus size-*B*-plus-color-*X*, and that these are not equivalent to the patterns size-*A*-plus-color-*Y* and size-*B*-plus-color-*Y*. More usually, however, the concept of stimulus-patterning is applied only when the identity or overlap of stimulus elements in the two discriminanda is more conspicuous. An example would be the discrimination of white-over-black versus black-over-white, or of the sequence of tones *c* — *e* from the sequence *e* — *c*. A difference in spatial or temporal arrangement of otherwise identical elements is characteristic of what is traditionally thought of as pattern discrimination. Such difference of spatial or temporal arrangement, however, is only a special case of stimulus patterning. The essence of the concept lies in the principle that any combination of stimuli may be discriminated from any other, regardless of the amount of overlap of elements, provided there is at least one point of difference. Thus, *ABCDEF* is discriminable from *ABCDEG* and from *ABDCEF*.

Let us see, now, how the peripheralist explains the solution of various discrimination problems by the mechanism of stimulus patterning, and without assuming that animals may respond to relational properties.

1. *Simultaneous Discrimination*

A stimulus-pattern may comprise either one or both of the discriminanda. In a simple two-choice discrimination, say of white + versus black —, for instance, the two objects or plaques may form an undifferentiated Gestalt or configuration. In conventional procedure, there are two such configurations, white-black and black-white. According to the theory elaborated by Gulliksen and Wolfle (3), the animal learns a left-going response to the first of these configurations, a right-going response to the second. The inapplicability of this movement-learning mechanism, in at least some cases, has been demonstrated (15).

Both Spence (20, 22) and the writer have described the learned responses in discrimination problems as approach to the positive stimulus or object, and nonapproach (Spence) or avoidance of the negative object. The only difference, here, between the peripheral and central interpretation is in the exclusion or admission of response to relative properties. According to the former, the discriminanda function only as absolute properties, with some generalization from one to the other. According to the latter view, the animal may approach the lighter or whiter of the two objects, and avoid the darker one.

2. *Successive Discrimination*

In one form of successive presentation, the animal responds actively to the positive stimulus and does not respond to the negative one (e.g., 2).[2] In other forms of the successive problem, the animal makes differential overt responses to the stimuli; it may, for instance, enter the left alley when the bifurcation is lighted, the right alley when the bifurcation is darkened (e.g., 26). Especially in the latter case, the movement-learning theory offers the simplest explanation: turn left for light, turn right for dark. In approach-nonapproach (or avoidance) terms, the determining stimuli include both light (versus dark) and the stimuli marking the left (versus right) alleys. There may, thus, be approach to the sensory patternings of light-plus-left and dark-plus-right, with avoidance of the other two stimulus combinations. An alternative centralist interpretation of this and the following problems will be described below.

---

[2] In an unpublished study of rhesus monkeys in these Laboratories, P. Hutt and M. Jarvik found that on appearance of the negative, nonrewarded stimulus, the animal retreated to the back of the cage, instead of just remaining passive. This occurred after some 20 trials; no punishment was used. This observation, of course, favors the concept of "avoidance" in contrast to that of "nonapproach."

3. *"Conditional Discrimination"*

I have placed this label in quotation marks, since it is here intended only as a descriptive designation of a problem situation, without reference to any explanatory concept. A simple example will serve to indicate this type of problem: the subject is trained to choose the larger of two white plaques (*W* vs. *w*), *and* the smaller of two black plaques (*b* vs. *B*). Two kinds of sensory patterning may serve as a mediating mechanism for solution. (*a*) There are four unique two-object configurations (*W-w*, *w-W*, *b-B*, *B-b*), and to each of these an appropriate left- or right-going response may be learned. (*b*) There are four unique one-object stimulus combinations or patterns, two of which (*W* and *b*) elicit approach, and two of which (*w* and *B*) elicit nonapproach. This second interpretation is presumably preferred by Spence, and is one which I consider possible but less plausible than the alternative discussed below. In discussing a "complex conditional reaction" based on the same principle but involving 16 habits using five cue-dimensions, I said, "it may be that each of the 32 plaques . . . provided a discrete, unique pattern of stimulation" (16, p. 14).

4. *"Transverse Patterning"*

The following problem was suggested by Spence (22) as involving "transverse patterning." The task of the animal is to master the three habits of Figure 1 so that they can be performed concurrently (i.e., correctly when presented in random order). The necessity of responsiveness to both objects on each trial is obvious: the triangle, for instance, is correct or incorrect depending on whether it is paired with circle or square.[3] Again, either of two types of sensory patterning may serve as the mediating mechanism. (*a*) According to the movement-learning formulation, the subject may learn left- and right-going responses to each of the six unique two-object configurations indicated in Figure 1. For chimpanzees, at least, this possibility has been disproven by the recent experiment of Thompson (24).

(*b*) Spence's "transverse pattern" mechanism has not been described in previous publications (cf., 22).[4] If a transverse pattern were merely an undifferentiated two-object configuration, selective approach-nonapproach would not be possible. Approach to a two-object pattern does not account for differential approach to one of those objects. Spence, however, postulates a pattern which comprises the perception of one form and the trace

---

[3] The solution of a similar problem, involving four instead of three habits, has been reported by Klüver (10). His squirrel monkey learned to choose yellow when paired with green, green when paired with red, red when paired with blue, and blue when paired with yellow. Each color was positive or negative depending on the color with which it was paired. The same principle was embodied also in a four-habit problem with chimpanzees, reported by Nissen (14, pp. 5-6).

[4] The description which follows is based on a personal communication from Dr. Spence (23).

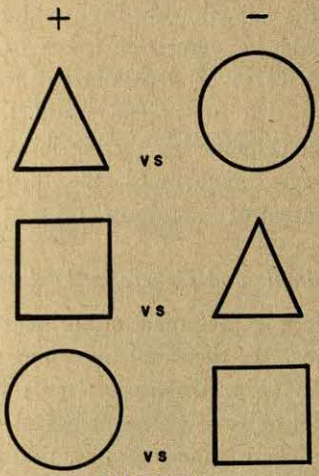


FIGURE 1
SPENCE'S "TRANSVERSE PATTERNING" PROBLEM: THE MIRROR IMAGE OF EACH PAIR SHOWN IS ALSO PRESENTED TO THE SUBJECT

of a perception of the other form. The subject looks first at one, then at the other; the second perception blends with the trace of the first perception.[5] If the perception of triangle, for instance, combines with the trace of a just previously perceived circle, the animal approaches. If, however, the perception of triangle combines with the trace of a just previously perceived square, the animal does not approach. Using capitals for perceptions, lower case letters for perception-traces, we have the following schema:

circle-TRIANGLE	R+	triangle-CIRCLE	R—
triangle-SQUARE	R+	square-TRIANGLE	R—
square-CIRCLE	R+	circle-SQUARE	R—

---

[5] This mechanism, or one very much like it, must be the basis for perception of a relation, assuming that there are such perceptions (see below). The perception of "larger," for instance, must derive from comparing two stimuli of different sizes, and must involve the specific temporal sequence: small-large. Perception of "smaller" derives from the same two stimuli in reverse order: large-small (cf., 11).

This fact—that a peripheralist should suggest a mechanism for one situation which is a necessary and integral part of the centralist position in respect to other behavioral data—may presage the eventual rapprochement of the two views. It should be pointed out, here, that the term "patterning" is or may be a very broad one. In its widest sense it is equivalent to "organization," since organization may be expressed in terms of spatial, electro-chemical or other patterning. To avoid such meaningless extension of the word, the present discussion uses the term, *sensory* or *peripheral* patterning.

As Spence points out (22), the temporal transverse patterning mechanism demands considerable precision of "preliminary receptor-exposure acts," and rather fine discrimination of perception-plus-trace patterns. If, as apparently is often the case, the animal looks back and forth at the two objects several times, there would be perception-traces of both objects, and the most recent one would have to be discriminated from just preceding ones. If it were to be shown that an animal could learn this problem with tachistoscopic presentation (eliminating the possibility of successive fixations) or with the head and eyes immobilized, the theory would then have to postulate a successive fixation of "attention" on first one and then the other figure.

5. *Simple and Generalized Matching*

The demand for precise sense-organ adjustments becomes even greater in other problem situations. In the simple matching situation (see Figure 2) as used with chimpanzees (17), the subject must look first at the sample, then at one of the choice objects. If sample and choice objects are alike, the animal approaches. There would thus be two trace-perception patterns to be approached (*aA*, *bB*), two to be avoided (*aB*, *bA*). Chimpanzees (17) and monkeys (25) learn this problem readily. Furthermore, they continue to match when new and unfamiliar objects are presented. The learning of specific patterns does not account for such generalization. Movement-learning to unique configurations is therefore ruled out as an explanation. However, as has been suggested (17, pp. 68-69, the "learning mechanism"), it is conceivable that such generalized matching is solved by one-trial learning with secondary reinforcement. (In original training the sample, as well as the matching choice-object, was rewarded. The position of the sample being constant, this may lend secondary reinforcement to the sample and to the choice-object which is like the sample.) On each trial the subject could learn, from its position, that the sample is correct, and could then transfer this learning to the identical choice-object.

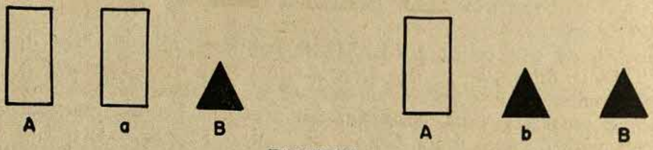


FIGURE 2
THE SIMPLE MATCHING PROBLEM
The sample is in the center, the choice objects on the sides. Mirror images of both 3-object settings shown are also presented.

6. *Conditional Matching*

In conditional matching (5, 18) the problem becomes more complicated (see Figure 3). In our example the animal must respond to the color-matching object when background *X* is used, to the form-matching object when background *Y* is used. There are now eight unique "transverse" patterns, each consisting of five essential elements (color and form of the sample and of the choice-objects, and the background) which lead to approach, and eight other patterns leading to avoidance or nonapproach. The adequate trace-plus-perception patterns are obtained only by looking first at the sample, then at the choice-object. If the glance (or attention) of the subject goes from one choice-object to the other one, or from a choice-object to the sample, irrelevant patterns (not leading to solution) are formed. Not only must the subject distinguish eight five-component patterns from eight other patterns, but, through a precise sequence of sense-organ adjustments it must eliminate the many other irrelevant and confusing trace-plus-perception patterns which the situation makes possible.

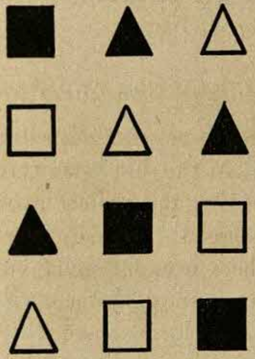


FIGURE 3

FOUR SETTINGS OF THE CONDITIONAL MATCHING PROBLEM

Sample in the center, choice objects on the sides. In these four settings the color-matching object (correct with X-background) is on the left, the form-matching object (correct with Y-background) is on the right. Mirror images give four more settings, and each of these eight is presented on both the X- and Y-backgrounds.

7. *Generalized Conditional Matching*

We have seen that peripheral patterning as explanation for conditional matching with a fixed set of stimulus-objects is not inconceivable. However, if the subject generalizes—continuing to choose the color-matching object under condition X and the form-matching object under condition Y when new colors and forms are introduced—then sensory patterning is no longer adequate as a mediating mechanism. Such generalization has been found, in monkeys, by Harlow (reported in 5 and 27). An example of the novel color- and form-combinations used follows: the sample is a red cross; the choice-objects are a green cross, red circle, yellow circle, and green triangle. Under condition X the red circle is correct, under condition Y the green cross is to be chosen. Solution of this problem requires something else than peripheral patterning, transverse or otherwise. What this something else may be is suggested in the following.

C. THE CONDITIONAL REACTION AND RELATIONAL PERCEPTION

The "conditional" problem $W+$ vs. $w-$ and $b+$ vs. $B-$ (see above) typifies a discrimination task which has been variously designated as a conditional reaction (12, 16), response to ambivalent cues or objects (14, 24), multiple-sign problems (4), and the "*if . . . then . . .*" problem (10). Although these several terms have been applied primarily in a descriptive sense, some of the discussions have implied that a special mediating mechanism is involved. Spence, however, considers the conditional problem as merely one instance of pattern discrimination, to be explained by the mechanism of peripheral patterning. Speaking of the successive discrimination problem, he says, "Instead of referring to it as involving a conditional reaction, however, the writer has referred to it as patterned discrimination" (22, p. 89). This attempted identification raises the problem of defining the two mechanisms and specifying the differences between them. Only then can their relative merits be evaluated. A crucial point is the question of whether a satisfactory explanation can be found for solution of the generalized conditional matching problem which, as we have seen, is not accounted for by the theory of sensory patterning in its present stage of development.

1. *Successive and Conditional Discrimination*

Figure 4 presents a highly simplified diagrammatic representation of the W vs. w, b vs. B problem according to (a) the pattern discrimination, and (b) the conditional or *if then* explanation.[6] In Figure 4a, a unique mechanism for each of the particular stimulus-combinations leading to approach is postulated: one for large-plus-white, another for small-plus-black. Activation of either center[7] requires double stimulation, e.g., from "large"

---

[6] By substituting symbols to represent *light, dark, left,* and *right,* these diagrams apply also to our example of the successive discrimination problem.

[7] As used in this paper, the word "center" has no implications for underlying

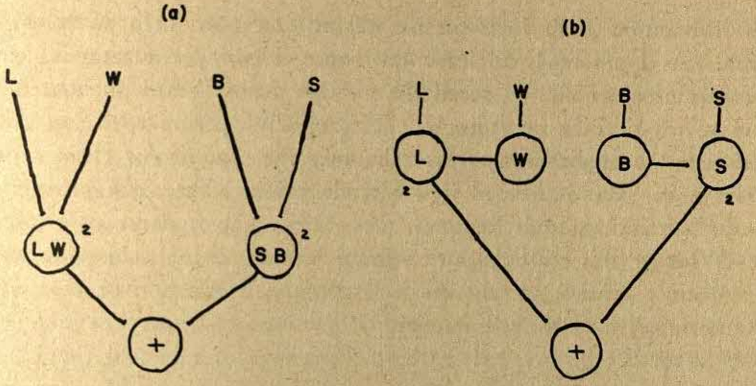


FIGURE 4

SIMPLE "CONDITIONAL DISCRIMINATION" ACCORDING TO (*a*) SENSORY PATTERNING, AND (*b*) *if then* FORMULATIONS

L, W, B, S refer, respectively, to large, white, black, small. The centers for stimulus patterns *LW* and *SB* in 4a, and the centers for *L* and *S* in 4b, require double stimulation in order to be activated.

and "white." (For the sake of simplicity in this and the following diagrams, centers leading to avoidance or nonapproach are omitted.)

In Figure 4b there are no special centers representing specific stimulus-combinations. Instead, it is assumed that the centers for "large" and "small," differentiating the objects presented for choice, lead directly to approach when facilitated by the appropriate aspect of the conditional stimulus-dimension (brightness). The stimulus "large" elicits approach when facilitated by "white," whereas "small" elicits approach when facilitated by "black." The conditional stimulus functions as a "set," a predisposition to respond in a certain way to another stimulus. *White* sets the switch for approaching *large; black* for approaching *small.*[8]

---

neurological structures or processes. A "center" is therefore a mechanism, for which a neurological basis is conceivable but unknown. The diagrams of Figures 4 through 7 are presented merely as concise and convenient symbolical representations of logical relations or "intervening variables."

In Figures 4, 5, 6, and 7, centers with a threshold or θ value of 2 (cf. 1) are designated by a small figure *2*. For all other centers θ is assumed to be 1.

[8] In the 16-habit conditional discrimination (16) mentioned above, there is a series of four conditions, or binary switchings, which determine response to the dimension (size) that differentiates the two choice-objects. There are eight sequences, of four successive alternatives each, eliciting approach to large, and eight other sequences eliciting approach to small (e.g., if white, if triangle, if margin, and if peg, then approach small).

Peripheral patterning, on the other hand, demands 16 one-object patterns, each consisting of specific combinations of five stimulus-aspects, which lead to approach. These must be discriminated from 16 other patterns differing from the first 16 in from one to five aspects (cf., 16, p. 10).

An illustration from introspective observation may help with the distinction. As I am ready to leave the house, I consider whether I should take my raincoat or not. I recall the weather report heard an hour before, and make my decision accordingly. This clearly appears to be an *if then* response: if the prediction was for rain, take the coat, if for clear weather, then leave it. But if instead it is already raining, the sight of rain plus that of the raincoat, leads "without thought" to the response of taking the coat. Whether this coalescence of stimuli into an unanalyzed pattern is in fact a distinct process, or whether it is illusory, a function of the smoothness and rapidity of a well-practiced *if then* process, may be questioned. Earlier in my life history, before the consequences of rain and the effectiveness of a coat were thoroughly known, such a decision may have involved a serial thinking through of the two *if . . . then . . .* alternatives.

The problem shown in Figure 1 and discussed above is schematized in Figure 5. Again, the patterning mechanism assumes a separate center for

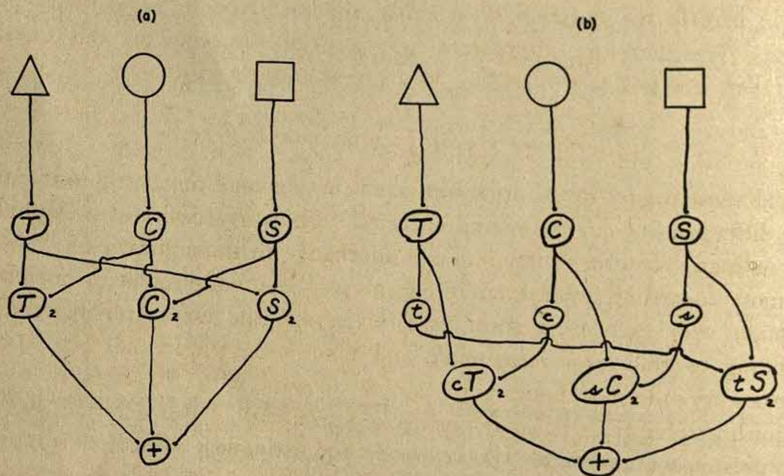


FIGURE 5
THE PROBLEM OF FIGURE 1 ACCORDING TO (*a*) THE *if then* AND (*b*) THE TRANSVERSE PATTERNING MECHANISMS

Upper case letters *T*, *C*, and *S* refer to perceptions of triangle, circle, and square, respectively; lower case letters designate traces of these perceptions. Centers requiring double stimulation are designated by the figure *2*; see Footnote 7.

In 5a, the triangle elicits approach only if there is facilitation from the circle. The circle and square likewise elicit approach only with facilitation from the square or triangle, respectively. In 5b, the approach response is elicited by *cT*, *sC* and *tS*, and these, in turn, are activated by two perceptions occurring in a specific temporal sequence.

each unique combination of sensory events—in this case, of serially ordered events. The *if then* mechanism, on the other hand, assumes that each of the three forms—triangle, circle, square—elicits an approach response when facilitated by the appropriate companion form.

It is clear that up to this point our two mechanisms or "hypothetical constructs" differ in form but not in effect. They are two ways of accounting for the behavior in question, the "intuitive properties" of which will carry varying degrees of conviction for different people (9). As long as problem-solution involves a fixed and invariable number of stimuli, the possibility of response to specific stimulus-combinations or patterns remains. Only when this possibility has been ruled out will there be a compelling need for the concept of the conditional stimulus as a predisposing set. A significant difference in explanatory adequacy will therefore appear only when we consider other discrimination problems, particularly that of generalized conditional matching. Intimately involved in these further problems, howeever, is one of the other subissues in the central-peripheral controversy mentioned above, namely the question of whether animals respond to relational properties. We therefore interpolate here a brief discussion of this question insofar as it refers to the relation of likeness and difference.

2. *Perception of Identity and Difference*

The sense organs of all organisms are being exposed constantly to identities and differences of environmental energies. The very essence of a "stimulus" lies in spatial heterogeneity or temporal change in those energies. It would be most surprising, therefore, if organisms were not provided, innately or through very early experience, with a mechanism for perceiving and responding to this most elementary and ubiquitous feature of their sensory world. We do not hesitate to endow animals with the capacity to distinguish greater from lesser intensities of stimulation. But does that capacity not presuppose and imply the capacity to distinguish some intensity from no intensity, and a change from a constancy of intensity? It is perhaps the very simplicity and obviousness of this fact which has led to its neglect. The English associationists (e.g., 13) accepted the capacity to perceive identity and difference as an ultimate fact, and built their systems of mental functioning on what seemed to them this self-evident and irreducible phenomenon. Since then, however, this elementary datum has been completely submerged in psychological theorizing. Although some of our leading theorists take for granted that animals perceive relations, as far as I know it forms no part of any current formalized system.[9] The postulation of centers for the perception of identity and difference does no violence to our knowledge of neural functioning. In this respect, indeed, it poses far less difficulty than does, for instance, the concept of primary stimulus generalization. Figure 6 presents a hypothetical mechanism for color identity and difference.

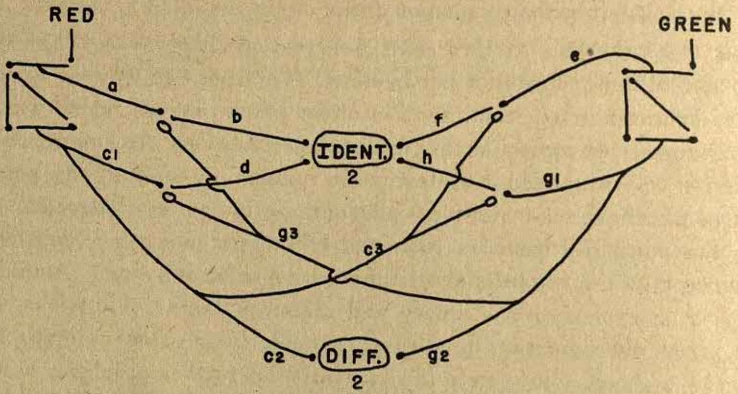


FIGURE 6

SUGGESTED MECHANISMS FOR PERCEPTION OF COLOR IDENTITY AND DIFFERENCE

Conventions of representation follow Culbertson (1). Solid dots indicate excitatory end-bulbs, hollow circles indicate inhibitory end-bulbs (dominant over excitatory endings). The triangular arrangement of neurons provides a reverberating circuit for each color.

Example: When two perceptions of red (or green) follow each other (i.e., when there is a temporally extended perception of one color), there is double stimulation of "Identity." There is stimulation from concurrent perception, via *a*, *b*, and from the corresponding "reverberating circuit" via *c1*, *d*. Activation of *c2* and *c3* is in this case without effect.

Example: When perception of green follows that of red, there is double stimulation of "Difference," via *c2* and *g2*. Stimulation of "Identity" by *d* (from the reverberating circuit of red) is inhibited by *g3*, and that from *f* and *h* is inhibited by *c3*.

3. *Matching—Simple, Conditional, and Generalized*

Accepting the perception of identity and difference as basic and universal in animal behavior, we have an explanation for simple and generalized matching which is at least as plausible, and somewhat simpler than the mechanisms which were discussed above. The subject perceives the identity of one choice-object with the sample and is trained, through differential rewarding, to choose it rather than the choice-object which differs from the sample. Since "identity" is not limited to any specific stimulus patterns, the subject continues to match when new objects are presented.

---

[9] One reason for this active rejection may lie in an exaggerated fear that, by readmitting any mechanism which guides behavior without benefit of learning, we lay ourselves open to a mass influx of faculties, soul, will, and instincts, of which we once rid ourselves so painfully. To this I would say that if we accept, as we must, the ontogenetic development of other extremely complicated and adaptive structures, such as the eye and its central connections, then the postulation of neural centers of perception should not strain our credulity too much. As Sherrington (19) has pointed out, there is no time or opportunity for trial and error, in the psychological sense, to guide the growth of the visual apparatus.

In accounting for "conditional matching," we may combine the concept of relational perception with that of the conditional reaction. As a possible aid in visualizing the process in crude outline, Figure 7 presents a sort of paradigm for conditional matching performance. In accordance with the foregoing discussion it is assumed that the animal is provided with mechanisms for the perception of form and color identity. Previous train-

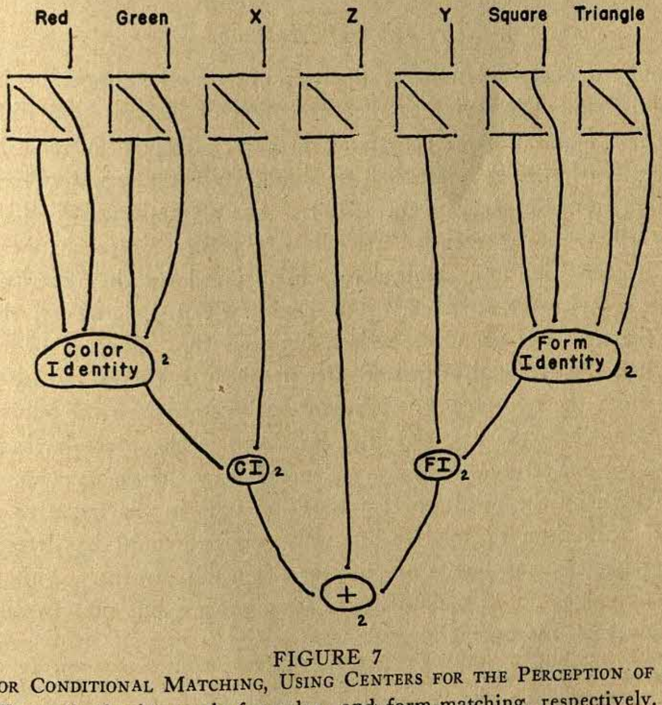


FIGURE 7

Schema for Conditional Matching, Using Centers for the Perception of Identity

X and *Y* are the backgrounds for color- and form-matching, respectively. *Z* indicates proprioceptive or other stimulation produced by shifting orientation from sample to choice object. It is assumed (*a*) that connections up to and including centers for Identity are innately or previously established, and (*b*) that there are similar connections for all colors and forms. The schema therefore permits generalization to new colors and forms. Associations centrifugal to (below) the Identity centers are established by specific training.

ing in simple matching has already accentuated responsiveness to this relation. Further training now differentiates the response (approach) to color identity when it is or is not facilitated by background *X*. Similarly, the subject learns to approach the form-matching object only when there is facilitation from *Y*. In both cases, overt approach requires facilitation from *Z*, representing the act of comparing the sample and choice-object (e.g., looking from one to the other). The *if then* or conditional reaction mechanism is represented in the figure by the second level centers, CI and FI. The perception of color identity is mediated by an established mechanism common to all colors, and thus differs from a particular sensory pattern acquired by specific training. Unlike the patterning mechanism discussed above, the present conceptualization permits generalization of the conditional matching habit, learned with one set of stimuli, to new colors and forms.

4. *Perception of More and Less*

Response to other relational properties of stimuli—almost as basic and universal in animal behavior as the perception of identity and difference—also has not found general acceptance as fact. Skepticism regarding relative perception has been reinforced, no doubt, by the alternative explanation of the data contributed by Spence (21). The attractiveness of his solution lies primarily in its promise of a single explanation or mechanism for the data on transposition and absolute response. Need for the "act of comparison" thus seems to be obviated.[10] It should be noted, however, that in removing the necessity of a comparison between the "objects" or their perceptions, the theory requires instead a comparison of the excitatory strengths deriving from the two objects. Without such comparison the subject would tend to respond to the first object it looked at, each object having a more-than-zero net excitatory strength—as demanded by the theory (pp. 434-435) and as shown by the results of transposition tests in the negative direction. Without comparison of relative strengths, choice would be determined by irrelevant habits or accidents of receptor-organ adjustment. This theory of transposition does not eliminate the relating process but merely displaces it, presumably to a more central region.

Another possibility, pointed out to me by Spence (23), is that by differential rewarding the positive stimulus gains a superthreshold effective excitatory potential $({}_{s}\bar{E}_{R})$, whereas that of the negative stimulus is below threshold. Choice is therefore determined by excitatory potential in relation to the threshold, rather than by the potentials of the positive and negative stimuli relative to each other. Absence of response to the negative stimulus (when looked at first), even though its potential may be greater than zero, is thus accounted for. This conceptualization, however, is inconsistent with Spence's theory of transposition. Let *A, B, C,* and *D* represent stimulus-plaques of increasing size (cf., 21, Figure 2). Suppose the animal is trained on *B*+ versus *C*—, and subsequently responds "relationally," i.e., chooses *A* as against *B,* and *C* against *D*. This result is explained on the basis of generalization gradients which give a greater net excitatory strength to *A* versus *B,* and to *C* versus *D*. The response is thus a function of the excitatory strengths of the two stimuli relative to each other. If the response were instead a function of absolute super- and subthreshold values established in training, the animal should approach *B* as readily as *A,* and should not approach C. (Analogous difficulties appear when the animal is trained to C+ versus B—.) At its present stage of development Spence's theory either implies an act of comparison or fails to account for transposition.[11]

---

[10] A very similar act, however, is reintroduced in the mechanism of "transverse patterning."

A central mechanism for the perception of "more" and "less," similar to that for identity and difference, would provide an explanation for the vast array of relational (transpositional) behavior displayed in all sense modalities by animals from ameba to man. It is neurally conceivable, being perhaps the most elementary case of the general problem of equivalence.[12] It may be a good working hypothesis to assume that, unless proven otherwise, behavior in a given situation is determined by learning in that situation. However, that hypothesis can only retard progress if it carries us to the extreme of blindly denying the evidence for mechanisms evolved phylogenetically or in the course of earlier experience.

D. SUMMARY

Peripheralist theories of behavior determination tend to emphasize the rôle of present stimulation and of current learning. The centralist theories, on the other hand, stress the importance of central structures, established innately or through prior learning, which organize the present sensory input.

---

[11] Dr. Spence (23) objects to this conclusion (which, it should be noted, does not and cannot presume to anticipate future modifications of his theory). He has described to me in correspondence a mechanism, involving the fatiguing of receptro-orienting acts, which he believes can overcome the difficulty. Especially since the mechanism is quite complicated, I must leave its formal exposition to Dr. Spence.

[12] None of the neural analogies for relative perception suggested thus far, however, is entirely satisfactory. See, for instance, Householder (7). Hebb (6, pp. 41-44), and Culbertson (1, pp. 32-46). We may, however, accept the "intervening variable" of relational perception without tying it to a particular "hypothetical construct."

These two approaches are here compared in their explanations of animal discrimination learning.

The concept of stimulus-patterning gives a reasonable explanation for solution of the simpler discrimination problems used with animals. In order to account for more difficult problems, however, the concept of patterning must be extended, to include the combining of temporally disparate events following each other in a certain order ("transverse patterning"). An experimental test of the trace-perception or transverse patterning hypothesis is suggested. The alternative *if then* mechanism offers an explanation for all such problems in terms of facilitation: a given stimulus elicits a certain response if, and only if, facilitated by another (conditional) stimulus. The conditional ("if") stimulus provides a set which determines the response to the second ("then") stimulus. The *if then* formulation seems to provide the simpler, more direct, and more inclusive explanation.

Applied to the conditional matching problem, the stimulus-patterning mechanism makes great demands on the subject in the way of precise sense-organ adjustments and discrimination of complex stimulus-combinations. Furthermore, this mechanism does not account for generalization to novel stimulus-aspects. The *if then* mechanism, in conjunction with a postulated perception of identity, does give an adequate explanation of generalized conditional matching. Response to identity and difference (of stimulus qualities and intensities) seems to be a primitive and universal phenomenon in animal behavior. It is therefore reasonable to assume that animals perceive these elementary relations. Neural analogues for such perceptual centers are less complicated than are the "nets" required for many other functions.

In the final section of the paper, it is suggested that postulation of mechanisms for the perception of more and less (i.e., of greater and lesser intensities, extents, etc.) provides a reasonable and straightforward explanation for the phenomenon of transposition, exhibited by all animals in all sense modalities.

REFERENCES

1. CULBERTSON, J. T. Consciousness and Behavior. Dubuque, Iowa: Brown, 1950.
2. GRICE, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. Comp. & Physiol. Psychol.*, 1949, **42**, 365-373.
3. GULLIKSEN, H., & WOLFLE, D. A. A theory of learning and transfer. *Psychometrika*, 1938, **3**, 127-149.
4. HARLOW, H. F. Responses by rhesus monkeys to stimuli having multiple-sign values. In Q. McNemar, & M. A. Merrill (Eds.), *Studies in Personality*. New York: McGraw-Hill, 1942.
5. ———. Solution by rhesus monkeys of a problem involving the Weigl principle using the matching-from-sample method. *J. Comp. Psychol.*, 1943, **36**, 217-227.

6. Hebb, D. O. Organization of Behavior. New York: Wiley, 1949.
7. Householder, A. S. Neural structure in perception and response. *Psychol. Rev.,* 1947, **54**, 169-176.
8. Hull, C. L. Principles of Behavior. New York: Appleton-Century, 1943.
9. Kendler, H. H. "What is learned?"—A theoretical blind alley. *Psychol. Rev.,* 1952, **59**, 269-277.
10. Klüver, H. Functional differences between the occipital and temporal lobes with special reference to the relations of behavior and extracerebral mechanisms. In *Cerebral Mechanisms in Behavior.* (The Hixon Symposium.) New York: Wiley, 1951.
11. Lashley, K. S. Studies of cerebral function in learning: VII. The relation between cerebral mass, learning, and retention. *J. Comp. Neurol.,* 1926, **41**, 1-58.
12. ————. Conditional reactions in the rat. *J. of Psychol.,* 1938, **6**, 311-324.
13. Mercier, C. A. Psychology, Normal and Morbid. New York: Macmillan, 1901.
14. Nissen, H. W. Ambivalent cues in discriminative behavior of chimpanzees. *J. of Psychol.,* 1942, **14**, 3-33.
15. ————. Description of the learned response in discrimination behavior. *Psychol. Rev.,* 1950, **57**, 121-131.
16. ————. Analysis of a complex conditional reaction in chimpanzee. *J. Comp. & Physiol. Psychol.,* 1951, **44**, 9-16.
17. Nissen, H. W., Blum, J. S., & Blum, R. A. Analysis of matching behavior in chimpanzee. *J. Comp. & Physiol. Psychol.,* 1948, **41**, 62-74.
18. ————. Conditional matching behavior in chimpanzee; implications for the comparative study of intelligence. *J. Comp. & Physiol. Psychol.,* 1949, **42**, 339-356.
19. Sherrington, C. S. Man on His Nature. New York: Macmillan, 1941.
20. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.,* 1936, **43**, 427-449.
21. ————. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.,* 1937, **44**, 430-444.
22. ————. The nature of the response in discrimination learning. *Psychol. Rev.,* 1952, **59**, 89-93.
23. ————. Personal communications, 1952.
24. Thompson, R. Approach-avoidance in an ambivalent object discrimination problem. *J. Exper. Psychol.,* 1953, **45**, 341-344.
25. Weinstein, B. Matching-from-sample by rhesus monkeys and by children. *J. Comp. Psychol.,* 1941, **31**, 195-213.
26. Weise, P., & Bitterman, M. E. Response-selection in discriminative learning. *Psychol. Rev.,* 1951, **58**, 185-195.
27. Young, M. L., & Harlow, H. F. Generalization by rhesus monkeys of a problem involving the Weigl principle using the oddity method. *J. Comp. Psychol.,* 1943, **36**, 201-216.

Yerkes Laboratories of Primate Biology, Inc.
Orange Park, Florida

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 289-294.

INDIVIDUAL DIFFERENCES IN SUBJECTIVE COLOR*

Cornell University Medical College and Hunter College

CHARLES A. KNEHR[1]

A. PROBLEM AND PROCEDURE

In the literature on subjective color, speculation centers around retinal functions or retinal modulation of optic nerve activity as the basic mechanism to account for the production of subjective color by intermittent retinal stimulation. An extensive review of the experimental work was published in 1949 (1).

A suspicion that the midbrain and calcarine cortex play more than a passive rôle in the development of subjective color perceptions has arisen as a result of an attempt to explain what appear to be changes in color perception in clinical patients following electro-convulsive treatment (ECT). While no consistency in effect could be found from one patient to another, in several patients blacks were consistently called purple or violet on a color naming test. One of the possible explanations seemed to be subjective color produced by unsteady fixation (5). Another possibility was that alpha activity might be altered centrally, and patterns of periodic impulses coming from the retina, when integrated with cortical activity, would then yield a different resultant or a changed perception.

In preliminary work, a disk was prepared using only the inner band of the Benham disk and rotated clockwise. At a speed of about three revolutions per second, three observers reported an olive green, two reported blue purple, and two reported a deep red with a bluish cast. Differences as marked as these do not appear in the literature, although one inconsistency between observers is known to have been attributed to refractive errors (1). The often reported warm-up period before the colors can be seen, in addition to these individual differences, may be considered as evidence against a solely retinal interpretation for subjective color produced in the normal way. To check the observed differences, the following experiment was performed.

A standard Benham disk was prepared and rotated clockwise against a medium gray background at 12.5 revolutions per second. The illumination

---

*Received in the Editorial Office on May 7, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Study from The New York Hospital and the Department of Psychiatry, Cornell University Medical College, New York, and Hunter College of the City of New York.

was tungsten light filtered through a MacBeth daylight screen with a reading of 4 fc. on the disk. This source was used so that subjects could make color matches with the 85 chips from the Farnsworth-Munsell 100-Hue test. Artificial pupils were used binocularly and measured .093 inches in diameter. Visual corrections, when worn, were placed in front of the artificial pupils after adjustment of the interpupillary distance. Twenty-five subjects were run. Eleven were males and 14 were females with ages ranging from 20 to 55 years.

The subjects were instructed to look at the rotating disk and, after a minute or two, to describe what was seen. If color was reported, the subject was further instructed to ignore saturation and brightness and to indicate its location in the spectrum by selecting one of the color caps most nearly matching. These caps are numbered and may be plotted on the color circle.

B. Results

The data for all subjects are plotted for each of the four bands separately, reading (and numbered) from the center of the disk out (Figure 1). The charts are those used for plotting results on the 100-Hue test.

The inner band yielded only two colors. Green was seen by two observers, while purple or blue purple was reported by six observers. The remaining 17 reported only gray, although some of these indicated warm grays, and others characterized the grays as cool.

The second band shows a wide distribution of color matching from yellowish-green around through yellow, orange, and red, to blue-purple.

The distribution for the third band is quite different with most observers reporting colors from the bluish green around to reddish orange.

In the fourth band, most observers reported charcoal or other dark gray with relatively few reporting color.

It was also found that, should two individuals report the same color in the first band, the second and third bands might not correspond. For example, KR and RN reported the inner band as purple (cap No. 65). For the former, the next two bands were reported as olive (No. 29) and purple (No. 64). The latter reported the second band as weak flickering orange (No. 10) and the third as the same but more yellow (no color cap match). Some subjects saw the second and third bands as differently colored while others saw about the same color but with slight to marked differences in saturation.

Repeat testing on six subjects indicated no appreciable changes in color matching. The removal of the artificial pupils at the end of the experiment did not alter the portion of the spectrum to which the observed color was assigned although differences in saturation were reported.

C. Discussion

In the literature on vision, measurements which appear to involve retinal functions yield results with the eye of one observer which tend to be highly similar for a second and third, provided the eyes are "normal." This has led to the apparent "elegance" (1) of some of the data and formulations in vision. Much of the data on subjective color has been reported on only a few observers, which may account for the failure to report extensive individual differences.

If subjective color were explainable on the basis of the differential rise and fall of the color receptors or by optic nerve modulation, some consistency would be expected from one subject to another under constant conditions of viewing such that a lawful statement could be made which would apply to several observers. The results indicate clearly that the effect is not comparable from one individual to another. Add to this the often observed need for a "warm up" period before color is perceived. The driving of the retinal mechanism must be so direct that this lag or "pulling in" (6) might be sought elsewhere. Nor is this delay similar for different observers. The additional fact that uniformity of gray in the bands at high rotation speed breaks down into different shadings at lower speeds must also be subject to the same explanation.

Recent work in electrophysiology and studies relating the electrical activity of the brain to photic stimulation indicates that retinal activity does not pass unchanged to the cortex. Walter and Walter (6), for example, attribute subjective visual effects to "interference between rhythmic evoked responses and spontaneous rhythms at cortical and possibly thalamic levels." Any incoming retinal data must be integrated into the excitability cycles of central neurone networks. Here, according to Lindsley (4) the most useful parameter in transforming and integrating data is time. Subjective color has previously been indicated to be due among other external variables, to differences in the pulsating character of the stimulus and its effect on receptors which respond differentially. Fry (3), among others, questions the ability of the three color theory to account for the phenomena. The large individual differences reported above also indicate that more than this is needed for explanation since one would expect different retinas to act similarly, or at least yield colors in the same order. Transformations in the

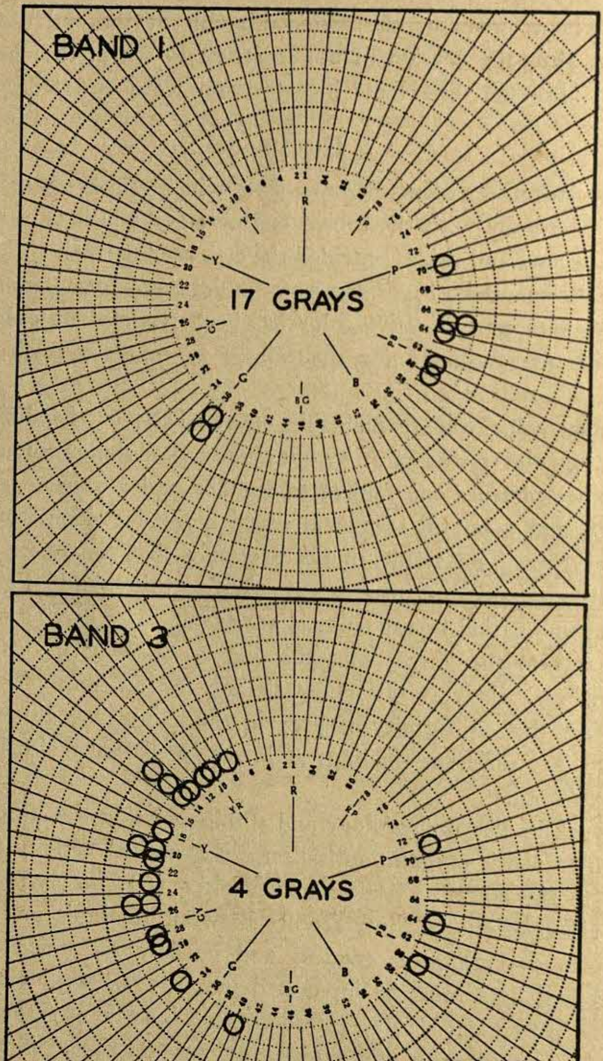


FIGURE 1
Distribution of Color Matches for the Four Bands of the Benham Disk Rotating at 12.5 Cycles/Sec.

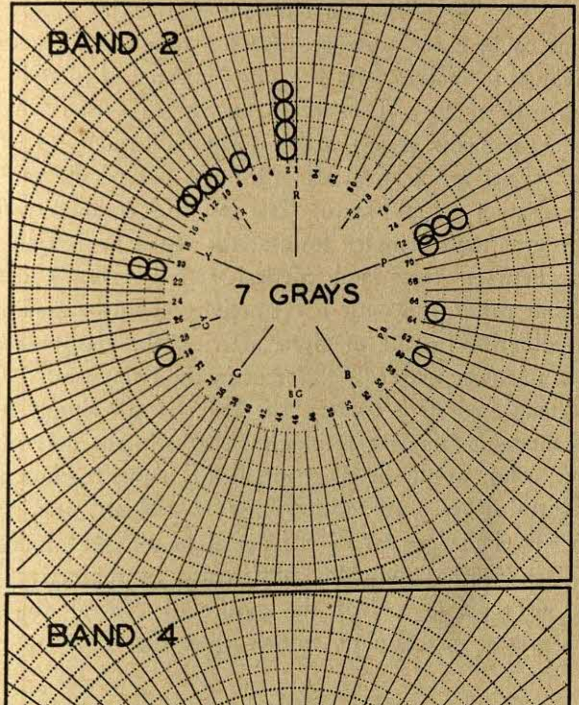


FIGURE 1
Distribution of Color Matches for the Four Bands of the Benham Disk Rotating at 12.5 Cycles/Sec.

nuclei and integration depending on different excitability cycles in different nervous systems could conceivably account for resulting effects which vary from one individual to another. As Fry (2) indicated, Troland's modulation theory assumes patterns to be transferred unchanged across synapses, an account which does not now accord with present indications. In the Benham disk, the on-off cycles are unequal and these are not of the same frequency and order of pulsation for all bands. The resulting complex retinal activity might arrive in even more complex form at the visual projection areas where time relations between these and cortical activity would determine the result. If similar patterns of optic nerve activity could be assumed to arrive in two brains differing in rate of alpha activity, the resultant activity and thus perception might be markedly different in the two cases.

If the foregoing analysis is correct, the brain is not simply a receiver of already coded messages from the retina, but shares in the determination such that subjective color would be a resultant of highly complex integrative activity of central locus. It follows that if retinal function and optic nerve activity are similar as far as the nuclei in different individuals, it is reasonable to assume that individual differences in subjective color might be indications of some aspects of the nature of central functioning. It would seem possible, furthermore, to put these assertions to experimental test.

D. Summary

Marked individual differences in reported subjective colors were observed under standard conditions. The possible rôle of central activity in subjective color has been discussed.

References

1. Cohen, J., & Gordon, D. A. The Prevost-Fechner-Benham subjective colors *Psychol. Bull.*, 1949, **46**, 97-136.
2. Fry, G. A. Modulation of the optic nerve-current as a basis for color-vision. *Amer. J. Psychol.*, 1933, **45**, 488-492.
3. ———. Color sensations produced by intermittent white light and the three-component theory of color-vision. *Amer. J. Psychol.*, 1935, **47**, 464-469.
4. Lindsley, D. B. Psychological phenomena and the electroencephalogram. *EEG Clin. Neurophysiol.*, 1952, **4**, 443-456.
5. Luckiesch, M., & Moss, F. K. A demonstrational test of vision. *Amer. J. Psychol.*, 1933, **45**, 135-139.
6. Walter, V. J., & Walter, W. G. The central effects of rhythmic sensory stimulation. *EEG Clin. Neurophysiol.*, 1949, **1**, 57-86.

Payne Whitney Psychiatric Clinic
525 East 68th Street
New York City 21

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 295-298.

SHIFTS IN THE VALUES OF OPINION ITEMS: FURTHER DATA*

Department of Psychology, Stanford University

JOSEPH-MAGDY AYAD AND PAUL R. FARNSWORTH

A. PROBLEM AND PROCEDURE

The belief that opinion items of the Thurstone type do not change their scale values with the passage of the years was put to test in 1940-41 (1). At that time the item weights of the Thurstone-Peterson "Attitude toward War" Scale, No. 34, Form *A* (6), were recalculated. One hundred eighty-three students of elementary psychology served as judges. The 1940-41 values thus obtained were found to differ appreciably from those found at Chicago in 1930-31 although the correlation between the older and newer sets of values was very high. The values of 13 of the 20 items showed shifts which were significant at the .01 level or better. Later, the opinions of a group of 256 students (not employed in the prejudging) were tested by the scale. These subjects were assessed first through the use of the original scale values and then by the newer ones. A critical ratio of 34.5 was found to obtain between the two group means. Hence, it seemed erroneous to assume that all item weights remain unchanged over the years or that the changes have no score significance.

The author of this 1940-41 study found it advisable in the description of his findings to state: "Yet it may well be that the item weights of other scales which are intended for the tapping of opinion in less dynamic areas may be found to be relatively stable." It was to follow up this latter suggestion that, in 1951, the current study was conceived. With the thought that religious ideologies might be more "family-bound" and so might not mirror so immediately changes in world conditions as did war ideologies it was decided to employ the following six scales.

- I. *Attitude toward God*
  - *A.* (The Reality of God)
    - 1. Scale No. 22, Form *A* (4)
    - 2. Scale No. 22, Form *B* (4)
  - *B.* (Influence on Conduct)
    - 1. Scale No. 22, Form *C* (3)
    - 2. Scale No. 22, Form *D* (3)

---

*Received in the Editorial Office on May 16, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

II. *Attitude toward the Bible*
   1. Scale No. 29, Form *A* (2)
   2. Scale No. 29, Form *B* (2)

The attempt was made faithfully to follow Thurstone procedures (5) in all details.[1] Sixty students from a class in elementary psychology at Stanford University, free to serve as subjects in any five of a considerable number of possible experiments, volunteered to be judges after being told the area of the proposed research.

TABLE 1
PERCENTAGES OF ITEMS OF SCALE 22 SHOWING SHIFTS IN VALUES

	Form *A*		Form *B*		Form *C*		Form *D*	
	% of Items	Sig. 1%	% of Items	Sig. 1%	% of Items	Sig. 1%	% of Items	Sig. 1%
Positive shift*	95	60	75	65	73	18	54	4.5
No shift	0		10		4.5		4.5	
Negative Shift	5	5	15	0	23	4.5	41	9

*The person who accepts an item whose value has changed in the *positive* direction receives on the new norms a more "religious" assessment than he would, had the 1931-33 values been employed.

TABLE 2
PERCENTAGES OF ITEMS OF SCALE 29 SHOWING SHIFTS IN VALUES

	Form *A*		Form *B*	
	% of Items	Sig. 1%	% of Items	Sig. 1%
Positive Shift*	86	36	91	36
No Shift	9		0	
Negative shift	4.5	0	9	0

*See footnote to Table 1.

TABLE 3
PERCENTAGES OF ITEMS SHOWING SHIFTS IN VALUES: TOTALS

	% of Items	Sig. 1%
Positive Shift*	79	36
No Shift	5	
Negative Shift	16	3
Total	100	39

*See footnote to Table 1.

B. RESULTS

Tables 1-3 show the percentages of the items which shifted in value between 1931-33 and 1951. They also show the percentages of the items whose shifts were significant at the .01 level or better. The overall value of 39 per cent is to be compared with 65, the percentage of significant shifts of the 1940-41 study of war attitudes. The conclusion seems inescapable that while there were important shifts in the semantic values of the religious items, these were less numerous, except in the case of Scale 22, Forms *A* and *B*, than was the case with the war items. This ideological area, in other words, appears to have been less mobile than that of war.

---

[1] Since Thurstone did not preserve his original variability data it was necessary to assume values identical to those obtained in the current recalculations.

It should be noted that while there were 36 per cent of significant positive changes in the six scales, taken as a group, there were only three per cent of negative shifts. Since a positive change means that the items in question are now regarded as closer to the "pro-religious" end of the continuum, another manner of describing the findings might be to state that the current ideologies appear to be slightly less "religious" than in 1931-33.

The need for employing item weights which are not out-of-date is shown by the data of Table 4. One hundred ninety-four students from general psychology classes at San Francisco City College were given the six opinion measures during the autumn of 1952. Their scores were figured both on the 1931-33 and on the 1951 value norms. Significant score shifts were found for the group for all scale forms except 22, Form *D*.

TABLE 4
Opinion Scores as Calculated from Two Sets of Weights

	Mean of Mds.-Old	σOld	Mean of Mds.-New	σNew	N	r	$D/\sigma D$
Scale 22, Form *A*	7.26	2.00	8.45	2.42	194	.98	28.48
Scale 22, Form *B*	6.76	2.60	8.05	3.14	194	.99	28.75
Scale 22, Form *C*	3.76	2.07	3.55	2.09	194	.99	9.32
Scale 22, Form *D*	3.57	2.28	3.55	2.20	194	.99	.67
Scale 29, Form *A*	7.67	1.83	8.25	2.00	194	.99	23.70
Scale 29, Form *B*	7.75	2.02	8.25	2.15	194	.99	26.35

C. Résumé

This study continues one done in 1940-41 in which the item values of a scale measuring attitude toward war were recalculated. In this earlier research 65 per cent of the items had value shifts over a 10-year period which were significant at the .01 level. In the present study the item values of six scales having to do with religious attitudes were recalculated. This general ideological area was chosen as being less dynamic, i.e., changes in religious attitudes should not parallel so promptly changes in world conditions. It was hypothesized that if this choice of a more static area were sound, the placement of religious items should be less disturbed by a passage of years than is the location of war items on a militarism-pacifism continuum.

After approximately 20 years the items of the two scale forms having to do with the reality of God were found to display value shifts of roughly the frequency previously found true for the war items which had been recalculated after 10 years had elapsed. The scale forms measuring attitudes toward the Bible showed far fewer value shifts of significance, while the scale values of the two tests purporting to measure the influence of God on conduct yielded still fewer significant shifts. The religious area appears, on the whole, to be less dynamic than that of war.

Twelve times as many of the significant (.01 level) item shifts were toward the "pro-religious" end of the continuum as were toward the "anti-religious." Such a finding lends support to the notion that the 1951 ideologies were slightly less religious than were those of 1931-33. Data were also gathered which show that persons currently scored or norms established two decades ago may not be properly assessed.

REFERENCES

1. FARNSWORTH, P. R. Shifts in the values of opinion items. *J. of Psychol.*, 1943, **16**, 125-128.
2. THURSTONE, L. L. A scale for the measurement of *Attitude toward the Bible;* Scale No. 29, Forms *A* and *B*. Chicago: Univ. Chicago Press, 1933.
3. ———. A scale for the measurement of *Attitude toward God* (Influence on Conduct); Scale No. 22, Forms *C* and *D*. Chicago: Univ. Chicago Press, 1931.
4. ———. A scale for the measurement of *Attitude toward God* (The Reality of God); Scale No. 22, Forms *A* and *B*. Chicago: Univ. Chicago Press, 1931.
5. THURSTONE, L. L., & CHAVE, E. J. The Measurement of Attitude. Chicago: Univ. Chicago Press, 1929.
6. THURSTON, L. L., & PETERSON, R. A Scale of *Attitude toward War*, Form *A*. Chicago: Univ. Chicago Press, 1931.

Department of Psychology
Stanford University
Stanford, California

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 299-309.

A FURTHER DEMONSTRATION OF PHI MOVEMENT CEREBRAL DOMINANCE*

School of Optometry, University of California

DARRELL B. CARTER[1]

A. THE PROBLEM

The experiment reported here was originally designed to ascertain whether a close relationship might not exist between ocular dominance determined by "retinal" rivalry, as reported in articles by Washburn, Faison and Scott (7); by Hamburger (2); by Küchle and Remky (5) and by others; and ocular dominance elicited by the Jasper phi movement test. The possibility that these two types of sensory ocular dominance are merely different demonstrations of the same phenomenon has been discussed by Walls (6).

The phenomenon of apparent movement visual dominance was first reported by Jasper (3). The experimental procedure consists of presenting alternately two luminous targets which are situated at different distances from the subject and lined up so that the one not fixated is seen diplopically equidistant both the right and left of the fixated target. When the two targets are alternately flashed at an optimal rate for the perception of movement, there is created an opportunity to perceive movement from the fixated center target either to the right or to the left, or to perceive the center target as splitting and going in both directions.

When movement is seen in one direction only and the other diplopic image is either suppressed or is seen as just blinking on and off (the report of all subjects in this experiment), it can be said that a visual dominance or laterality is present which may be of either a peripheral (ocular) or of a central (cerebral) character. For example, with binocular fixation on the far target, the perception of movement from the center (far) target to the right diplopic image of the near target is attributable either to a dominance of the left eye or left half of the cortex. With near fixation, the perception of movement from the fixated target to the right diplopic image of the far

---

*Received in the Editorial Office on May 20, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The author wishes to express appreciation to Professor Gordon L. Walls for suggesting this experiment and to Professor Mark R. Rosenzweig for advice and guidance in the preparation of this paper. The author also wishes to thank the American Optometric Foundation for their fellowship stipend.

target is the result of a dominance of either the right eye or the left hemisphere; movement left conversely indicates left eye or right hemispheric dominance.

By combining the near and far fixation results, a subject can be classified into one of the four categories of left cerebral, right cerebral, left ocular or right ocular dominances; or, of course, as uncertain or unreliable. It will be immediately noted that the same subject cannot demonstrate both a cerebral and an ocular dominance.

Jasper reported upon 77 normal subjects and 16 stutterers. Of the 31 right handed normals (to convert Jasper's percentages back to the actual numbers), four subjects were right eyed, five left eyed, 15 left cerebral, five right cerebral, and two completely ambilateral. His 22 left handed normal subjects showed the following dominances: four right eyed, six left eyed, none left cerebral, nine right cerebral, and three ambilateral. The 24 non-stuttering ambidextrous subjects divided into seven right eyed, three left eyed, three left cerebral, three right cerebral, and eight ambilaterals. The 16 stutterers (whose handedness was not given) show one right eyed, four left eyed, none left cerebral, one right cerebral, and 10 completely ambilateral.

The high incidence of left cerebral dominance in right handed normals; of right cerebral dominance in left handed normals; and the equality of occurrence of cerebral dominance in non-stuttering ambidextrous subjects clearly indicates a close relationship between the brain laterality that results in handedness and that which is the basis of the phi movement cerebral dominance phenomenon. The occurrence of cerebral dominance in only one of 16 subjects among the stuttering group also indicates a relationship may exist in many cases between stuttering and the absence of the cerebral dominance that is demonstrated by the Jasper phi movement test. As Jasper apparently ran his subjects only once, he reported no data as to the reliability of his results.

Jasper and Raney (4) repeated this two-target phi movement test on 42 boys between the ages of 9 and 12 years. Of these, on the first run, 30 were left cerebral, four right cerebral, six right eyed, one left eyed, and one indeterminate. On the second run, 32 were left cerebral, four right cerebral, four right eyed, one left eyed, and the same indeterminate individual. The results of one subject were not tabulated because he was unable to perceive motion. The test-retest reliability was given as 90 per cent. In reporting the papers of Jasper and of Jasper and Raney, I have converted their terms "left central dominance" and "right central dominance" to, respectively, right cerebral and left cerebral dominance so as to indicate the actual cerebral hemisphere involved, rather than the direction of the perceived movement.

B. APPARATUS

The apparatus consisted of a compartmentalized light-tight box which was painted black; the subject looked through two small diamond-shaped openings into the upper part of the apparatus box. The upper compartment contained the targets which differed from those used in the Jasper phi movement apparatus in three main particulars.

First, they were not dots but instead were erect rectangles, the far target being 2 mm. by 7 mm. and the near target 1 mm. by 3.5 mm. These sizes were selected so that the two targets would have the same apparent size, and were situated so that they were aligned with the mid-point between the subject's eyes, the distances from the subject being 82 cm. for the distant target and 41 cm. for the near one.

Second, a small, apparently self-luminous red fixation dot situated just above the target was provided for both the far and near targets. The red fixation dots also allowed the subject to realize quickly when he had lost binocular fixation.

Third, the targets were given a green hue by inserting a wratten No. 54 filter between them and the incandescent bulbs which provided the illumination. This was done to eliminate a color temperature difference that existed between the near and far targets after they had been equated for brightness.

A major addition to the Jasper apparatus was made by providing two more targets of the same size and color and the same distance from the observer as the distant target, situated 2⅞ inches to either side of the distant target. The use of these targets will be designated by the term three-target situation (3TS). The modified Jasper apparatus will be called the two-target situation (2TS).

The targets were produced from lucite rods by masking off with black friction tape all but the target area. The lower ends of the lucite rods were inserted into the lower half of the box which contained the incandescent lamps. The intensity of each target was individually controllable by means of a variable transformer. The flashing mechanism consisted of two cams which opened and closed the two snap action switches which controlled the lights. The rate of alternation of the lights was changeable over a wide range by a differential pulley and belt system which connected an electric motor to the cams. This arrangement did not permit a continuous variation of the flashing rate, but this was not necessary as it was quite easy to find a rate of alternation that gave good apparent movement. With nearly all subjects, a flash time for each light of 1/10 of a second combined with a dark interval of 4/10 of a second between flashes (thus making one second the total period for one cycle) was found to give a perception of apparent motion within a few seconds.

C. PROCEDURE

1. *Two-Target Situation*

The procedure, always carried out in a darkened room, consisted of first having the subject experience apparent movement for about 10 minutes outside the experimental apparatus by means of two lucite rods which were illuminated alternately. The subject was then seated at the apparatus, with either the far or near red fixation dot turned on, and was instructed to fixate it. The corresponding target was turned on and was of course seen singly and just below the fixation dot. Then the other target was turned on and was seen in either crossed or uncrossed diplopia (depending on whether the far or near target, respectively, is fixated) with one diplopic image on either side of the fixated target. The subject adjusted the head rest until the two diplopic images were judged to be equidistant from the fixated target, and the height of the three images appeared the same. In the 2TS, the two targets to either side of the distance target were never on and played no part in the 2TS.

The subjects were instructed to report the character of any apparent movement that they observed. They were told to maintain a passive attitude and not to try to see any particular movement. The flashing mechanism was then started. Nearly always the subject would first report movement in both directions from the fixated target. That is, the fixated target would be seen splitting and moving both to the left and to the right and then both would move back to the center and fuse. After a period varying from four or five seconds to as long as four minutes, but usually about 30 seconds, the subject would report that movement was perceived on only one side of the fixation target, the center rectangle apparently moving to that side and then back to the center. The other diplopic image was then perceived as just flashing on and off. No subject failed to perceive movement; some subjects in some runs continued to see movement in both directions for an entire run of four minutes. Two subjects never saw movement in one direction only, always seeing movement in both directions simultaneously.

2. *Three-Target Situation*

In the three-target situation the near target and its corresponding fixation dot were removed from the apparatus. The same instructions were given to the subject. The 3TS was done binocularly and with each eye separately.

The subjects in the 3TS again would nearly always first perceive the center target as splitting and moving in both directions to the two side targets and then those two would move back together and fuse. After a variable period, nearly all subjects would again report that the center target was moving in only one direction, left or right, and the other target was just blinking off and on.

The order in which the five tests of 2TS-near fixation, 2TS-far fixation, 3TS-binocularly, 3TS-left eye only, 3TS-right eye only, were performed by the subject was varied randomly. Usually the subject made only three runs on any one day. While up to four runs were made by some subjects using one particular test situation, only the first two runs were used for the purpose of this experiment.

In recording the reports of subjects, the following procedure was used. If, during a run, a definite movement to one side only was reported and this movement continued for one minute without any change in character, the run was immediately ended and that trial was listed as showing a strong right or strong left result. This is indicated in Table 1 as *R* or *L*. On some runs, some subjects would report that the character of the movement fluctuated at about 10- to 30-second intervals, i.e., in both directions, now right only, now both, now right only, etc. Such a report was classified as a slight right report, and is designated by *r*. A slight left report is designated by *l*. The asterisk indicates that the subject was unable to maintain fixation.

Runs in which movement in both directions was seen continuously were considered as indicating ambilaterality and are designated by *A*. One subject (Subject 5) during three runs reported a definite left movement alternating with a definite right movement. This result is also designated by *A*.

The final classification of the subjects as to their results on the 2TS and the 3TS was made by essentially averaging the results of the various runs in each situation. The symbol used in the table indicates the direction of movement. The cerebral dominance is, of course, just opposite. For example, *R* indicates strong right movement and therefore strong left cerebral dominance.

Handedness for the purpose of this experiment was determined by the subject's own evaluation of his handedness. As the subjects used were all adults of above average intelligence, this method was considered to be accurate.

TABLE 1
RESULTS OF 27 SUBJECTS ON THE TWO- AND THREE-TARGET SITUATIONS: THE MEANING OF SYMBOLS USED IN THE TABLE ARE INDICATED IN THE TEXT

	Results of Trials 1 and 2					Classification	
Subjects Right handed	2TS Near fixation	2TS Far fixation	3ST Binocular	3TS Left eye	3TS Right eye	2TS	3TS
1	R-R	R-R	R-r	R-R	R-R	R	R
2	R-l	R-l	R-R	R-R	R-R	A	R
3	A-R	A-R	R-R	R-R	R-R	r	R
4	R-R	R-R	R-R	R-R	R-R	R	R
5	A-A	A-A	A-A	A-A	A-A	A	A
6	r-r	*-*	R-r	R-R	R-R	*	R
7	R-R	R-R	R-L	L-L	L-R	R	A
8	R-R	R-R	R-R	R-R	R-R	R	R
9	R-L	L-L	L-L	L-A	L-R	A	A
10	R-R	R-R	R-R	R-R	R-R	R	R
11	*-*	r-R	R-R	R-r	R-R	*	R
12	R-R	R-R	R-R	R-R	R-R	R	R
13	R-R	A-R	A-R	R-R	R-R	R	R
14	R-R	R-R	R-R	R-R	R-R	R	R
15	R-R	R-R	R-R	R-R	R-R	R	R
16	R-R	R-R	R-A	A-A	A-A	R	A
17	R	A	R	A	r	r	r
18	R	R	R	R	R	R	R
19	R	R	R	R	R	R	R
Left handed							
20	A-l	A-l	L-1	L-L	L-L	l	L
21	L-L	L-L	L-L	L-L	L-L	L	L
22	*-*	L-*	L-L	L-L	L-L	*	L
23	L-L	L-L	R-R	R-R	R-R	L	R
24	A	A	A	A	A	A	A
25	A	A	A	A	A	A	A
26	L	L	L			L	
Ambidextrous							
27	R-R	R-R	R-R	R-r	R-r	R	r

The experimental results reported in this paper are those obtained after the apparatus was modified. The subjects, nearly all students at the University of California School of Optometry, were tested on both the 2TS and 3TS. They had been screened to eliminate persons who did not have single binocular vision with at least 20/30 acuity in both eyes with correction lenses if needed. They were unaware of the nature of the experiment or the apparatus, although they had heard that the experiment concerned "eye dominance."

D. RESULTS

As previously stated, this experiment was undertaken to look for persons who might manifest a sensory ocular dominance by the Jasper phi movement test, so as to correlate the laterality of this supposed type of ocular dominance with that shown in "retinal" rivalry experiments. After screening 38 persons with the unmodified Jasper apparatus, the author had found only four subjects whose results on the Jasper apparatus indicated an ocular dominance.

In these screening runs, no continuous fixation points were provided, as both Jasper and Jasper and Raney had stated that their subjects had little or no difficulty in maintaining fixation. It was soon discovered that several subjects were unable to keep binocular fixation on the target to be fixated (for example, the far one) when it went off and the other one came on. Several others could maintain fixation only with great difficulty. This was true even though most of my subjects had had considerable experience with visual experiments which required accurate binocular fixation.

The Jasper apparatus was therefore modified at this point by adding the red fixation dots previously described. In addition, the 3TS was devised and added to the apparatus.

Once a continuous fixation point was provided, three of the four subjects who had seemed to have an ocular dominance no longer demonstrated it, but instead showed a cerebral dominance in accord with (that is, opposite to) their handedness. All three of these subjects reported great difficulty in maintaining binocular fixation of the target at the particular fixation distance where a false report would have resulted in their previously indicated ocular dominance. The fourth subject was not available for re-testing after the apparatus was changed to its final form.

The test-retest agreement in both the 2TS and the 3TS was determined from the first and second runs of each subject, using only three categories—left movement; right movement; and ambilateral. In discussing variability of results the author has lumped the near and far fixation distances for the 2TS; and the binocular, right and left eye findings for the 3TS situation as the experimental results indicate that each of these two groups are determined by the same factors.

Of 38 test-retest situations in the 2TS there was agreement in 30 cases, left-ambilateral or right-ambilateral disagreements in three cases, and left-right disagreement in five cases. Of the 63 test-retest situations in the 3TS there was agreement in 57 cases, left-ambilateral or right-ambilateral disagreement in three cases, and left-right disagreement in three cases.

Of 19 right handed subjects in the 2TS, 12 were left cerebral; two slightly left cerebral; three ambilateral, and two could not maintain fixation even with the red fixation dots present. One of the ambilaterals (Subject 9) indicated some right cerebral tendencies. On the 3TS, 14 were left cerebral, one slightly left cerebral, and four ambilateral. The one subject classified as ambidextrous—as she writes left handed but does all other common tasks with the right hand—was left cerebral on both the 2TS and the 3TS. Of seven left handed subjects in the 2TS, three were right cerebral, one slightly right cerebral, two ambilateral, and one could not maintain fixation. In the 3TS, of six subjects, three subjects were right cerebral, two ambilateral, and one left cerebral.

E. Discussion

This last subject (No. 23) gave results that indicated that the factors determining the laterality exhibited in the 2TS may be in some way different than those of the 3TS. Moreover, Subject 7 gave somewhat similar results. Possibly these results are the product of the difference in difficulty of observation between the 2TS and the 3TS, rather than possessing physiological significance. Or perhaps the 2TS is more influenced by the handedness of the subject. It is, of course, difficult to decide how much of the obviously close relationship that exists between handedness and the subject's report on this phi movement test is the result of an influence of handedness upon the mental set of the observer and how much of this relationship is produced by the dependence of both of these factors (handedness and the phi movement reports) upon a physiological cerebral dominance.

Three pieces of evidence bear on the intensity of the dominance demonstrated by the phi movement lateral dominance test. First, with a subject who reports, for example, right movement, it is possible to increase the brightness of the center and left targets in the 3TS, and obtain a report of what effect this has upon the perceived movement. Subjects who have a strong cerebral dominance continue to perceive their usual type of movement until there is an extreme difference in brightness between the center target and the target to and from which they normally perceive movement.

Second, a person with a moderately strong phi movement laterality can perceive the motion in the opposite direction only with a great deal of "mental effort" and practice. Immediately upon ceasing the attempt to shift the direction of movement, it switches back to the subject's usual perceived direction.

The third piece of evidence occurred because some of the subjects had had lectures, as part of their optometry studies, on possible differences between the nasal and temporal halves of the human retina. After completing the monocular 3TS tests, they expressed surprise that the movement had been in the same direction with both eyes, as they had expected to find a nasal-temporal difference.

Often the subject would report that the diplopic image that was just flashing on and off appeared after the movement outward on the other side was completed. This implies that speed of perception may be a factor in determining in which direction movement is perceived. The reader will realize that any difference in the perception time between the two lateral images, in the two-target situation, must be due to the observer, as they are really the diplopic images from the two eyes of the same physical target.

The results of Jasper and of Jasper and Raney which showed the existence in some subjects of a phi movement ocular dominance are probably attributable to their failure in both experiments to provide fixation points, as well as any failure in eliminating persons who are functionally "one-eyed." The failure of accurate fixation allows the image of the supposedly fixated target to slip off the fovea of one or both eyes. Since the heterophoria of a subject will usually be of such a character that he can maintain fixation on one of the two targets but not the other, it can be easily seen that inaccurate fixation may result in an incorrect report for only one of the two fixation distances.

Such a combination of one correct result (believing as I now do that the direction of perceived movement is determined by the dominant hemisphere) and an incorrect one will give a result that would be interpreted by Jasper and by Jasper and Raney as an ocular dominance. Incorrect reports at both fixation distances are much less likely and thus there will be few reports of a brain dominance of opposite laterality from the true one. Also, it should be noted that in the 2TS when loss of binocular fixation has occurred without either the subject or the experimenter being aware of the fixation loss, that the subject's report will depend upon which of the eyes continues to maintain fixation. If, for example, with near fixation, the right eye continues to fixate correctly, the near target will be seen as moving to the right. If the left eye holds fixation, the target will be seen as moving left.

The three-target situation has three main advantages over the two-target situation. First, it is much easier for subjects to maintain fixation. However, even with the 3TS, a constant fixation point is desirable.

Secondly, the existence of any condition which renders a subject primarily or completely one-eyed will cause a report in the two-target situation that will be interpreted as an ocular dominance. Such a report in cases where one eye is blind, is suppressed, or is highly amblyopic adds nothing or almost nothing to the experimenter's understanding of the subject's condition and adds nothing to an understanding of the ocular dominances that exist with many visual functions (i.e., sighting dominance and "retinal" rivalry dominance) in persons with normal single binocular vision and with approximately equal visual acuity in the two eyes. In dealing with "one-eyed" subjects, the three-target situation possesses the advantage that the cerebral laterality of these subjects can still be ascertained. A possible disadvantage of the 3TS arises in that it does not permit the demonstration of a perceptual ocular dominance if this phenomenon should really exist in an occasional individual.

The third advantage of the 3TS is that it enables monocular use of the phi movement test. Ammons and Weitz (1) have indicated phi movement is perceived more easily monocularly than binocularly, and that the monocular phi perception is more stable than binocular. One reason for this may be that monocular viewing eliminates any difficulty caused by the effort required for binocular fixation. As most phi movement experiments, including the one reported here, provide little or no clues for peripheral fusion, maintaining binocular fixation will be difficult for any subject with a significant heterophoria.

F. SUMMARY

1. The Jasper phi movement test for cerebral and ocular dominance was repeated on 27 subjects using a modified version of the Jasper apparatus. It is concluded that phi movement ocular dominance does not exist in persons with single binocular vision and approximately equal acuity in both eyes.

2. A new technique (the three-target phi movement test) is described which permits a determination of perceptual cerebral dominance monocularly as well as binocularly. This technique was used on 26 subjects. The advantages of the three-target test are discussed.

3. The relationship between phi movement cerebral dominance as determined by both of the tests and handedness is given for each subject. Consideration is given to the influence of mental set upon the results of both of the tests.

REFERENCES

1. AMMONS, C. H., & WEITZ, J. Central and peripheral factors in the phi phenomenon. *J. Exper. Psychol.*, 1951, **42**, 323-332.
2. HAMBURGER, F. A. Der Wettstreit und seine Rolle im Binokularsehen. *Kl. Monatsbl. f. Augenheilk.*, 1949, **115**, 289-319.

3. JASPER, H. H. A laboratory study of diagnostic indices of bilateral neuromuscular organization in stutterers and normal speakers. *Psychol. Monog.*, 1932, **43**, 72-174.

4. JASPER, H. H., & RANEY, E. T. The phi test of lateral dominance. *Amer. J. Psychol.*, 1937, **49**, 450-457.

5. KÜCHLE, H. J., & REMKY, H. Über monoculare Dominanz ("Äugigkeit"); alte und neue (haploskopische) Methoden zu ihrem Nachweis und deren Wert für ihre Beurteilung. *Arch. f. Ophth.*, 1951, **152**, 62-88.

6. WALLS, G. L. A theory of ocular dominance. *Arch. of Ophth.*, 1951, **45**, 387-412.

7. WASHBURN, M. F., FAISON, C., & SCOTT, R. A comparison between the Miles A-B-C method and retinal rivalry as tests of ocular dominance. *Amer. J Psychol.*, 1934, **46**, 633-636.

School of Optometry
University of California
Berkeley 4, California

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 311-322.

CREATIVITY AND CULTURE*

Department of Psychology, University of Chicago

MORRIS I. STEIN

In this paper[1] a series of hypotheses will be discussed regarding the personality of the creative individual, his work, the process through which he achieves it, and some of the relationships between these and the culture in which they appear. These hypotheses were developed in the course of studying the personalities of a small number of chicago artists.[2] The tentative nature of the hypotheses should be emphasized—their validity is still in question. They are now being subjected to test in a study of chemists.

Let us start with a definition. The creative work is a novel work that is accepted as tenable or useful or satisfying by a group in some point in time. Each of the parts of this definition will be considered separately.

By "novel" I mean that the creative product did not exist previously in precisely the same form. It arises from a *reintegration* of already existing materials or knowledge, but when it is completed it contains elements that are new. The extent to which a work is novel depends on the extent to which it deviates from the traditional or the status quo. This may well depend on the nature of the problem that is attacked, the fund of knowledge or experience that exists in the field at the time, and the characteristics of the creative individual and those of the individuals with whom he is communicating.

Often, in studying creativity, we tend to restrict ourselves to a study of the genius because the "distance" between what he has done and what has existed is quite marked. Such an approach causes us to overlook a necessary distinction between the creative product and the creative experience. The child who fixes the bell on his tricycle for the first time may go through stages that are structurally similar to those which characterize the work of the genius. His finished product, however, is a return to a previously existing state of affairs. The product of an inventor's labor,

---

*Received in the Editorial Office on May 26, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Adapted from a paper presented at the Northwestern University Centennial Conference on Anthropology and Psychiatry, May, 1951.

[2] The study of artists was supported by a grant in 1950 from the Social Science Research Committee of the University of Chicago. The study of chemists is now supported by the Research Division of Armour and Co., Chicago, Illinois.

on the other hand, may strike one as creative immediately because it did not exist previously. In speaking of creativity, therefore, it is necessary to distinguish between internal and external frames of reference.

Turning to the characteristics of the creative experience or creative process and the personality of the creative individual, we adopt a bipolar point of view, in which there is an *interaction* between the creative individual and the problem on which he is working, or, in broader terms, and the environment in which he exists. To speak solely of the existence of the stresses and strains in the environment without due consideration of the individual, as some investigators do, or to deal primarily with the stresses and strains in the individual and to overlook the nature of the problem or the environment as other investigators do, is an arbitrary approach which is a consequence of the specialization in our profession today. Such separate emphases, however, can yield only partial insight and understanding.

The first question that arises in analyzing the creative process is the question of motivation: "Why does the individual create?" This does not differ from any other motivational problem. Therefore, in the early stages of the creative process, the individual experiences a state of disequilibrium—one might say that homeostasis is disturbed, or that there is a lack of closure, or, from a hedonistic point of view, that the individual experiences a lack of satisfaction with the existing state of affairs. On probing more deeply into the roots of the individual's personality, one may realize the historical factors and personal needs which determine the subject's *sensitivity* to such states. The creative person has a lower threshold, or greater sensitivity, for the gaps or the lack of closure that exist in the environment. The sensitivity to these gaps in any one case may stem largely from forces in the environment or from forces in the individual.

Associated with this sensitivity is the creative individual's capacity to *tolerate ambiguity* (1). By "capacity to tolerate ambiguity" I mean that the individual is capable of existing amidst a state of affairs in which he does not comprehend all that is going on, but he continues to effect resolution despite the present lack of homeostasis.

To summarize this early stage of the creative process: the creative individual may be characterized as a system in tension sensitive to the gaps in his experience and capable of maintaining this state of affairs. Some individuals go no further than this point in the creative process. Their creativity is manifested in the fact that they have played a critical rôle in calling the attention of others to the gaps that exist. But for others the creative process continues to the next stages—hypothesis formation and hypothesis testing. These individuals seek various solutions that would close the gap or that would effect closure.

To be capable of developing such hypotheses, it is suggested that just as there need be some communication between the individual and his environment (i.e., that which was termed sensitivity previously), so there need be communication between some or all of the inner personal regions. Stated somewhat differently, the creative individual is characterized by permeable boundaries that separate the self from the environment and that separate some or all of the regions within the self. At times when this permeability does not exist, it may be induced by the taking of drugs or alcohol as we find to be the case in many creative persons. Or it may exist when the person is distracted or devoting himself to other works.

The character of the inner personal region obviously varies with the nature of the work that is undertaken. For persons in one area (physics, for example) it may mean greater flexibility in the intellectual sphere, while for others, the artist, it appears as a greater flexibility in the emotional or affective sphere. To be sure, there is an interaction between the two spheres, and rigidity in one area may well impede developments in the other and sidetrack the creative process.

This interaction and some of its hazards are well illustrated in Schiller's response to a friend who complains of his lack of creative power. He says:

> The reason for your complaint lies, it seems to me, in the constraint which your intellect imposes upon your imagination. Here I will make an observation, and illustrate it by an allegory. Apparently, it is not good—and indeed it hinders the creative work of the mind—if the intellect examines too closely the ideas already pouring in, as it were, at the gates. Regarded in isolation, an idea may be quite insignificant, and venturesome in the extreme, but it may acquire importance from an idea which follows it; perhaps, in a certain collocation with other ideas, which may seem equally absurd, it may be capable of furnishing a very serviceable link. The intellect cannot judge all those ideas unless it can retain them until it has considered them in connection with these other ideas. In the case of a creative mind, it seems to me, the intellect has withdrawn its watchers from the gates, and the ideas rush in pell-mell, and only then does it review and inspect the multitude. You worthy critics, or whatever you may call yourselves, are ashamed or afraid of the momentary and passing madness which is found in all real creators, the longer or shorter duration of which distinguishes the thinking artist from the dreamer. Hence your complaints of unfruitfulness, for you reject too soon and discriminate too severely (2, p. 193).

The "momentary and passing madness" is, I believe, a function of the permeability of the boundaries in the inner personal regions that are usually blocked from consciousness. The intensity of the "momentary and passing madness" is correlated with the depth to which the experience goes. For the artist it may go deeper into subjective experience than for the scientist. Instances of such experiences are described by many creative persons. Thus Zervos, in discussing Picasso's creative process, says:

> His only wish has been desperately to be himself, in fact he acts according to suggestions which come to him from beyond his own limits. He sees descending upon him a superior order of exigencies; he has a very clear impression that something compels him imperiously to empty his spirit of all that he has only just discovered, even before he has been able to control it, so that he can admit other suggestions. Hence his torturing doubts. But this anguish is not a misfortune for Picasso. It is just this which enables him to break down all his barriers, having the field of the possible free to him, and opening up to him the perspectives of the unknown (4, p. 109).

For some persons the creative process may stop during the stage of hypothesis formation. They develop too few hypotheses because of intellectual or emotional reasons, or in the process they become sidetracked by considering a specific intellectual matter that is not relevant to the demands of the moment, or they come upon a previously unresolved emotional difficulty that forces its attention upon them.

The process of hypothesis formation in the creative person is not a haphazard nor rigid process. It is a flexible one that is often characterized by either implicit or explicit direction. The creative individual's time perspective is oriented toward the future. He senses in the present how some aspects of the final form of the product are to appear. Thus Wertheimer, in discussing Einstein's thought processes while he was working on the theory of relativity, says:

> Before the discovery that the crucial point, the solution, lay in the concept of time, more particularly in that of simultaneity, axiom played no rôle in the thought process—of this Einstein is sure. (The very moment he saw the gap, and realized the relevance of simultaneity, he knew this to be the crucial point for the solution.) But even afterward, the final five weeks, it was not the axioms that came first. "No really productive man thinks in such a paper fashion," said Einstein.

Later he added:

> During all those years there was a feeling of direction, of going straight toward something concrete. It is, of course, very hard to express that feeling in words; but it was decidedly the case, and clearly

> to be distinguished from later considerations about the rational form of the solution. Of course, behind such a direction there is always something logical; but I have it in a kind of survey, in a way visually (6, p. 183).

✓ It has been observed by some that in the course of the creative process the creative individual experiences depression. The hypothesis is suggested that this depression arises as a result of anxiety that is brought forward by the lack of direction just mentioned. The creative person no longer feels that he is going forward and still he cannot enjoy the present state. The lack of direction may be a consequence of an excessive number of hypotheses that occur to the subject and the feeling of inadequacy as to the possibilities of testing any of them. It may also arise as a result of his inability to communicate his ideas to others. Finally, as the result of the relaxation of the inner personal barriers, old unresolved tensions are brought to the surface which are the residues of earlier life experiences wherein the individual was realistically inadequate—thus there is a process of reinforcement. Indeed, for the creative individual this is only one point in a sequence of events, but for others it may be the end.

After the development of a series of hypotheses, or even simultaneously with them, there is testing of the hypotheses. The testing may vary with the area of the work. For one person it may involve the construction of practical models, while for others it may involve changing the features of a painting. In any case, it is suggested that when the final solution is attained, that is, when there is closure for the individual, he experiences a feeling of satisfaction with the final work, a feeling of exhilaration with the good gestalt. It is this feeling that is manifest in "Eureka" or "This is it!"

Some investigators in the area of creativity who have relied primarily on the Rorschach test have remarked, with some degree of dismay if not disturbance, that the records of their subjects contain responses that are to be found in the records of severely disturbed persons. These responses apparently occur in sufficiently large numbers so that the emotional stability of the subjects is questioned. I would like to suggest that this finding may result from any one or a combination of the following factors. It may indeed be an accurate reflection of the creative person's instability, and it is this instability which may result in the subject's being sensitive to the lack of closure in his environment. It was therefore suggested previously that in the inner personal regions there is communication between *some* or *all* of the regions. The creative person may not necessarily be completely integrated. Some of the factors that differentiate him from the neurotic or the more severely disturbed person, such as tolerance of ambiguity, direction, and a time perspective oriented to the future, have already been discussed. He is also different from the neurotic or psychotic in the next stage—that of communication, which will be discussed below. Furthermore, the Rorschach finding may be a manifestation of the extent to which the creative individual differs from others, and is congruent with his tendency to perceive accepted reality in a manner that differs from most people. By deviating from the traditional and status quo, the individual has achieved the first stage in arriving at the novel work. Furthermore, since many of the so-called "bad" signs in the Rorschach are inadequately formed *Gestalten,* his responses may reflect his ability or capacity to tolerate ambiguity in one area—the perceptual area. The Rorschach responses are then only a sample of an individual's behavior in one area, from which we may infer his behavior in other areas. Finally, to fixate completely on the disturbed aspects of a creative individual's personality is to emphasize a finding that may be an artifact of our present theory, which is oriented, possibly much too heavily, toward psychopathology, and an artifact of the experimental design, which does not take into account aspects of the personality for which the Rorschach is not necessarily the best. Mind you, it is not suggested that the Rorschach finding is invalid, but that it is only part of the story. The part that is not considered sufficiently is the creative person's ability to convey or to communicate his personal experiences to others so that they may react to it.

This brings us to the second part of the definition suggested above: that the creative work is tenable or useful or satisfying. These terms were selected to cover the areas of ideas, things, and aesthetic experiences respectively, although they may well be replaced by other terms. My essential purpose here is to develop the thesis that the results of the creative process must be communicated to others. Communication with the self alone is insufficient. The creative person must achieve, as Sullivan says in another context (5), "consensual validation." This may be clearer if we take the case of the psychotic person. Much of what the psychotic does has significance only within a narrow idiosyncratic framework. At times he may not even be able to communicate his experiences adequately, as one finds in the "word salads" and neologisms of the schizophrenic. And further, the ideas he has do not stand up under test. But the creative person is able to convey his experiences so that they can be tested, or, if he has tested them, so that they may be reacted to by others.

Suggesting that the creative work may be regarded as an element in communication implies at least two factors: (*a*) The creative person must have available to him means or media through which he can express himself. Just as each of us required training in language to convey our ideas, so it is obvious that to be creative in painting the artist needs some experience in the use of pigments, charcoal, etc., or to be creative in mathematics, and not in others. Such a study might well reveal the congruence that exists between the personalities of the subjects and the areas in which they have selected to work. (*b*) In the course of communicating his ideas, the creative person needs to abstract or eliminate certain aspects of his work which are completely of himself. In the final stages of communication, the individual must, as Mead would say, bear in mind the "others" with whom he is communicating. The extent to which this occurs varies with the area of work. Thus it is more prevalent in science than in art. But in either case there is a process of evaluation in which completely autistic factors, or the difficulties that were experienced in the course of arriving at a solution are eliminated. Thus in Einstein's work:

> The way the two triple sets of axioms are contrasted in the Einstein-Infeld book is not at all the way things happened in the process of actual thinking. This was merely a later formulation of the subject matter, just a question of how the thing could afterwards best be written. The axioms express essentials in a condensed form. Once one has found such things one enjoys formulating them in that way; but in this process they did not grow out of any manipulation of axioms (6, p. 183).

Similarly, we find that Zervos reports on Picasso's work:

> I see for others, that is to say that I can put on canvas the sudden apparitions which force themselves on me. I don't know in advance what I am going to put on the canvas, any more than I decide in advance what colors to use. Whilst I work, I take no stock of what I am painting on the canvas. Every time I begin a picture, I feel as though I were throwing myself into the void. I never know if I shall fall on my feet again. It is only later that I begin to evaluate more exactly the result of my work (4, p. 109).

It should be indicated that at times the creative person may present his work so well after his evaluation that others regard it as "simple," and wonder why they never thought of it.

The next major portion in the definition of the creative work is acceptance by a group. It is suggested that the creative product is congruent with the needs or experiences of a group. It strikes a chord for the group as it does for the individual. The creative product "resonates" with the needs or experience of a group. The use of this word with regard to creativity reminds one of the resonance theory of hearing, in which it is postulated that different wave lengths of sound are heard because they strike different parts of the basilar membrane. To make an analogy between this and creativity, one might say that creative products resonate all along a responsive membrane. Thus certain art works resonate with feelings, while inventions resonate because they fulfill practical needs. In the case of art, it should be pointed out that there is not always a one-to-one relationship between what the artist attempted to express and what it resonated in the group. The conditions under which this does or does not occur is a matter for further investigation.

Acceptance by a group is significant. It provides the creative worker with his final test of reality, if you wish. The size of the group may vary. At times it may be only those on the "left bank," Greenwich Village, or the small number of persons who first gathered around Freud and Einstein. Individuals may seek out the creative person or he may have to proselytize them. The group provides the individual with necessary feedback so that he can clarify, alter, or make progress in his future work.

Finally, we arrive at the last point in our definition that the creative work is accepted at some point in time. The historical point of view reflected by the phrase "some point in time" is inserted in the definition to account for the fact that some men, like Van Gogh, may not be considered creative in their own lifetime; as a matter of fact, some, like Socrates, pay dearly for their ideas. Yet when the works of art or the ideas are rediscovered or are brought to attention at some future date, when the forces of society have changed, it is only then that these persons attain their rightful places in history. Indeed in some areas of creativity there may be "universals," e.g., form and use of space in art with only content varying in time. But in determining the validity of such universals, one needs to be aware of the hazard that what we regard as the critical elements in the universals today is a function of ourselves, and it may not necessarily be congruent with the manner in which these works were regarded in their own times.

The discussion, up to this point, has been limited primarily to the creative individual and the creative process. Let us now turn to a consideration of the interaction between the processes described and culture.

If what was said previously regarding the sensitivity of the creative individual in relation to both his external and internal environments is valid, then it may be said that a culture fosters creativity to the extent that it provides an individual with the opportunity to experience its many facets. A culture that limits the freedom of a person to study in one or a variety of areas cuts down his opportunity to pick out the gaps that exist in the culture and also keeps him from learning the necessary media of communicating his feelings or ideas.

A culture also fosters creativity to the extent that its parent-child relationships and child-rearing techniques do or do not result in the setting up of rigid boundaries in the inner personal regions.[3] Techniques that result in excessive repression or guilt restrict internal freedom and interfere with the process of hypothesis formation. Attention must also be directed toward the broader aspects of education. For example, does the culture tolerate deviation from the traditional, the status quo, or does it insist upon conformity, whether in politics, science, or at school? Does the culture permit the individual to seek new experiences on his own, or do the bearers of culture (parents, teachers, and so on) "spoon-feed" the young so that they constantly find ready-made solutions available to them as they come upon a situation that is lacking in closure? Furthermore, to what extent do the adults accept or reward and thus reinforce the creative experience that the individual has had? For example, in the case I spoke of earlier—the child who fixed his tricycle bell—his experience could have been handled either by a depreciation of his experience and verbalized as, "Oh, anyone could have done that!" or the magnitude of his experience for himself could have been recognized and he could have been encouraged to seek similar experiences in the future. Experiences of this kind should be studied both in the home and in formal educational systems.

The stage of development of a culture obviously influences the means available to the individual for creative progress. Thus the modern physicist has new vistas open to him as a result of the recent developments in nuclear physics. From the experience in this area, it may be said that the variety of creative works that occur in a culture vary with the number of works that deviate markedly from the traditional and are accepted. Only with the acceptance of the theory of relativity and other findings was the present development of physics possible, and we may well expect many new discoveries and applications in this area.

The culture may be marked by strict adherence to a specific philosophy and this too influences progress in the arts and sciences. Giedion (3) suggests how a mechanistic view of the universe has effected the work in many areas of inquiry. And, in the course of modern history, we have seen how various political and religious movements have at times stimulated and then markedly limited progress. Points of view which are developed in one area and studied by a worker in another way may aid the latter in sensitizing him to the gaps in his fund of knowledge. This is manifest in the recent developments in communication theory in the social sciences which were stimulated by the physicists. The extent to which progress is made in the new area of application may well depend on the rigidities of the philosophy that is adopted and/or on the manner in which it is interpreted and followed by those who accept it. For example, present-day mechanism and materialism has caused Giedion to suggest that we have been overlooking humanism and he highlights the need for a man who can live in "equipoise."

---

[3] Indeed, certain cultures set up rigid boundaries in certain regions and not others. This will no doubt affect the areas of creative work.

Furthermore, philosophies of life undergo cyclic change as a result of a multitude of factors. When a "valley" appears, the extent to which a culture comes out of it depends on the extent to which it is capable of tolerating ambiguity and encourages and tolerates a diversity of viewpoints as a new philosophy is developed. Further research might therefore be centered around the comparison of creative works when there is a consistent *Weltanschauung* and when one is lacking or when there is a conflict of philosophies.

Finally we come to a specific aspect of the culture—the audience with whom the creative person communicates—the critic, the patron, the followers, and the population at large.

The critic plays a very significant sociological rôle in determining what the larger population has available to it as instances of creativity. When we started our study of creative artists, we had a difficult time in getting critics to agree on who was a creative painter. Reputable persons in the art world disagreed with each other. Now let us assume that one of the critics, as was the case, is the curator of a gallery. That person would hang the works of one artist and not the other. For some people, because of the prestige of a museum, this work now becomes a "creative" work. But what of the others? They have to wait to be "discovered" by other critics and this may even happen, if they are fortunate, in their own lifetime. It might be wise, therefore, to study the personalities of critics or the supporters of scientific research, since they play such important rôles in determining what is creativity.

In some respects similar to both the critic and the supporter of scientific research, is the patron of arts. He, by virtue of the fact that he contributes financially to the artist's support, must be communicated with. How much of the patron's needs enter into the artist's evaluation of his final work depends on the relationship that exists between the two. The patron, just as was said of the broader culture, may well direct and restrict the content of his artist's work, if not the manner in which he works, because of his financial position. This raises a more crucial question in our own time when there are few, if any, patrons of the arts who are comparable to the Medicis, let us say. Is there anything to the hypothesis that *one* of the factors resulting in the development of non-objective art is the fact that the artist no longer is subservient to a patron and he can therefore express himself as he wishes?

In considering the final stages of the creative process, it was suggested that the creative work must strike a chord or resonate in some manner with the group that accepts it. If this hypothesis is valid, then it would follow that the personalities or the experiences of the group are in some manner similar to those of the creative individual. It may be further suggested that studies of creative products provide some data for making inferences regarding the needs of the group that are being satisfied. Thus it may be suggested that one of the factors involved in the acceptance of modern art today is that it represents a retreat or a rebellion, if you wish, against the materialism and intellectualism that has marked much of our time. It is a voice speaking against a period in which there is a denial of certain feelings. To be sure, to make similar interpretations of previous cultures may be hazardous if this is accepted as the only evidence. We cannot completely put ourselves in the place of primitive man.

Finally one must also consider the problem of communication between the creative individual and the population at large. In some areas of works, such as practical inventions, communication may exist with an extremely large portion of the population because they are aware of an invention's demonstrated usefulness. Indeed communication may stop when it is necessary to understand the thinking that has gone into the development. Modern artists have been most vociferous in asking why their art is not appreciated more widely than it is, especially when they as individuals feel that they have captured the spirit and problems of the day. This raises the question whether in an "age of anxiety" some individuals do not find it safer for themselves not to be aware of the problems that exist around them. Just as society affects the creative process by developing individuals who cannot relax the boundaries in the internal and external regions, so the extent to which an audience does resonate with an art product is a reflection of the extent to which they as perceivers are capable of relaxing their defenses.

Thus by studying the "non-appreciators" we can also learn a great deal of the culture.

In summary then: A definition of the creative work was presented. A creative work is a novel work that is accepted as tenable or useful or satisfying by a group in time. In line with this, it was hypothesized that studies of the personality of the creative individual may reveal a sensitivity to the gaps that exist in his own culture and his creativity may be manifest in calling attention to these gaps or in finding a means of effecting closure, or his sensitivity to certain facets of this may result in his desire to communicate to others. In addition to sensitivity, attention was called to the creative individual's ability to tolerate ambiguity and to maintain direction as he develops and tests his hypotheses. The final product, considered as an element in communication, resonates with the needs or experiences of a group at some point in time. In considering cultural factors, it was hypothesized that the extent to which a variety of creative products are developed depends on the extent to which cultural influences permit the development of both freedom between the individual and his environment and freedom within the individual; on the extent to which the culture encourages diversity and tolerates the seeming ambiguity that such diversity suggests. We also considered certain aspects of the audience, specifically the critic, the patron, the appreciators, and the population at large with whom the creative person communicates. The experimental problems involved in testing the hypotheses mentioned are numerous, but they are not insurmountable.

References

1. FRENKEL-BRUNSWIK, E. Intolerance of ambiguity as an emotional and perceptual personality variable. In *Perception and Personality* (Bruner, J. S., & Krech, D., *Eds.*) Durham: Duke Univ. Press, 1949.
2. FREUD, S. The Basic Writings of Sigmund Freud (Brill, A. A., Ed.). New York: Random House, 1938.
3. GIEDION, S. Mechanization Takes Command. New York: Oxford Univ. Press, 1948.
4. READ, H. Art Now. New York: Pitman, 1948.
5. SULLIVAN, H. S. Conceptions of Modern Psychiatry. Reprinted from *Psychiatry,* 1940, **3**, and 1945, **8**.
6. WERTHEIMER, M. Productive Thinking. New York: Harper, 1945.

Department of Psychology
University of Chicago
Chicago 37, Illinois

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 323-340.

CLINICAL DIAGNOSIS OF READING DISABILITY*

Educational Clinics, Queens College and City College of New York

ALBERT J. HARRIS AND FLORENCE G. ROSWELL

In child guidance clinic practice, children with reading disabilities usually comprise a large percentage of the total case load. Except in those cases in which the disability is considered of minor significance in comparison to other problems presented by the child, the clinical psychologist usually carries the major burden of the diagnostic study. There is a vast literature on the nature, causation, and treatment of reading disabilities (4), but very little has been written on the planning and conduct of a diagnostic psychological examination in this very important area of clinical practice.

A. DIAGNOSTIC ISSUES

Children with reading disabilities usually do not come to the attention of the clinical psychologist until several years after the difficulty has become apparent. By that time the clues are old and unreliable, and the task of reconstructing the early phases of the learning situation so as to be able to judge what was responsible for poor performance has become difficult or impossible. In many cases of reading disability, the number of handicaps or possible causal factors is sufficiently large and the historical information obtained from school, parents, and child is sufficiently vague to make an exact explanation of causation impossible. Nevertheless, the careful clinician will not be satisfied with an appraisal of current status. He will try to discover each significant factor that has a bearing, and to determine (as exactly as circumstances permit) its chronology, strength or severity, causal importance, and amenability to treatment.

Medical factors are often important. Defects of vision are common, particularly those involving eye fatigue and difficulty in achieving and maintaining binocular fusion. School results indicating normal vision are untrustworthy in the case of reading problems because they tend to ignore those factors which are likely to interfere with good vision at reading distance. Poor hearing, while infrequent in reading disabilities, may be highly significant when it occurs. Endocrine dysfunction may have an important bearing,

---

*Received in the Editorial Office on May 29, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

particularly when delayed sexual maturation is part of the picture along with thyroid and pituitary insufficiency. The debilitating effects of chronic poor health require careful evaluation; malnutrition, rheumatic fever, asthma, and tuberculosis are not uncommon, and any condition resulting in lowered energy may interfere with efficiency in concentration and effort.

Neurological complications are more common than is recognized in much of the psychological and educational literature. Since the organic deviation is usually slight and of the type which Gesell and Amatruda (14) have called "minimal brain injury," the orthodox neurological examination is usually negative, even when the case history and the psychological findings on such tests as the Rorschach, Goldstein-Scheerer (16), and Bender (2) are strongly suggestive of organic impairment. In our experience, the results of electro-encephalography are also quite often inconclusive when the psychological signs such as organic rigidity, concreteness and perplexity are quite clear. There is frequently a history of prolonged and difficult labor with marked moulding of the skull, early sucking difficulty, after-effects of oxygen deprivation, or infantile convulsions. A history of prematurity must be regarded with suspicion, since approximately one premature infant out of 10 shows neurological complications (9). Delayed speech development is frequent in these cases with organic complications. There is sometimes evidence of a defective rote memory. On performance tests there may be a pattern typical of organic deficit, or the difficulty may be confined to linguistic development and allow a high level of non-verbal functioning.

On psychometric tests such as the Stanford-Binet or Wechsler tests wide scatter is frequent, so that the conventional measures of mental level and *IQ* are often poor indicators of capacity for learning to read. We do not consider poor reading to be a special disability, however, unless the child's average mental level is at least a year higher than his reading achievement level. As one would expect when the causation is so varied, there is no consistent patterning in psychometric tests. Inferior memory may be found, particularly in auditory rote memory and in visual memory. In an individual case, one is often puzzled as to whether this is a genuine defect or a temporary symptom of anxiety. There are wide variations in vocabulary, comprehension, grasp of logical relations, problem-solving, etc. The clinician should be aware that *IQ*'s obtained from group tests in school are often greatly in error, since inability to read the questions prevents the child with reading difficulty from showing intelligence on such tests. Peculiarities in visual perception and organization may show up in drawing and copying tests. The Bender (2) is particularly useful for this purpose.

Despite the fact that children with reading disabilities are handicapped in acquiring verbal skills, many of them show much less competence in handling non-verbal than verbal tasks. The Kohs Block Design Test (27) and its various abbreviated versions are particularly sensitive to the kind of inability to analyze and synthesize visual patterns that is common in reading disabilities. Digit-symbol substitution tests reveal the slowness of associative learning that is another major characteristic. For these reasons, we feel that it is highly desirable to include a performance scale. We prefer the *WISC* Performance (39) or the Cornell-Coxe (6) to the other available performance scales for testing children with reading handicaps.

The relation of directional confusion to language disturbances is still a highly controversial subject, in which the available evidence is inconclusive. For years, many clinicians who have worked with speech, reading, and writing disturbances have been impressed by the frequency of deviations from a consistent right-sided pattern. Several theories to account for such a relationship have been advanced, of which the most widely discussed has been that of Orton (34), which postulates a conflict between left and right cerebral hemispheres as the basis of both the incomplete lateral dominance and the disturbances in language. This and other theories of lateral dominance are still unproven hypotheses. We believe that the important question is not whether there is a deviation from right-sidedness, but rather whether or not there is a real directional confusion. To check on this possibility, a careful qualitative exploration of hand, eye, and foot dominance is desirable. For this, we employ the Harris Tests of Lateral Dominance (20).

Careful inspection of the cumulative school records sometimes reveals information of major importance such as absence at critical times, interrupted schooling, frequent changes of school or teacher, the time when poor work first became evident, and the relationship of achievement to personality and conduct ratings, etc. Recollections about early school experiences are often highly significant. Which teachers were liked or disliked and for what reasons; what special help has been received and under what circumstances, promotion or non-promotion,—all these may throw light on the development of the problem. A common pattern is that of the child who does poorly in first grade, is promoted into second grade and expected to do second grade work, and from then on never makes an adjustment.

It is quite unusual to find a case of reading disability with no emotional difficulties, and one of the most important diagnostic issues is the determination of the nature and severity of personality disturbance. Closely related to it is the question of whether or not the child is likely to adjust well to remedial teaching or is more likely to require psychotherapy before remediation or concurrent with it. In some cases, response to remedial help is enthusiastic and motivation is sustained at a high level over long periods of time. Some maintain an attitude of indifference for quite a while to protect themselves from the renewed failure that they anticipate, but gradually thaw out and develop real interest when they find themselves making progress. For some, the reading disability is a way of clinging to infantile dependence and refusing to grow up; their outward compliance masks a strong resistance, and they may show little or no long-term progress in spite of apparent day-to-day learning. This type of neurotic immaturity is sometimes very difficult to distinguish from a real developmental immaturity. Extreme withdrawal or rigidity may make it impossible for the child to learn. Overt dislike may create real refusal to coöperate, or anxiety may be so great as to inhibit intellectual functioning in a reading situation. Occasionally one may find a case in which printed material has phobic significance or in which really traumatic experiences have caused severe negative conditioning to reading. Gauging the child's accessibility to remedial teaching involves careful weighing of the data obtained from the history, observations, interviews, tests, and sample lessons.

The child with a reading disability does not have any consistent pattern of cultural setting, family position, or emotional relationships. He may be oldest, youngest, middle, or an only child. There may be parental overprotection, rejection, or ambivalence. Sibling rivalry problems are frequent and a more successful sibling may be quite a discouraging factor. A fairly common pattern is that of the oldest child whose first sibling is born when he is in kindergarten, with resultant emotional upset at a particularly critical time for school adjustment. The average child with a brilliant older sibling is another recurrent pattern. Most children referred as reading disabilities tend to be inhibited and repressed in emotional expression; but the full range is from apathy to extreme excitability, and from complete conformity to open delinquency.

We have become increasingly aware of the fact that many children who fail to learn to read are children whose growth patterns are atypical. As emphasized by Olson (33), average growth in size and general mental development may conceal a striking physiological and anatomical immaturity, accompanied by marked delay in readiness for reading instruction. Readiness for formal learning may not develop in these children until around the age of eight, nine, or ten years. By that time, frustrations and pressures at school and at home have usually developed emotional reactions intense enough to be mistaken for causal factors.

Related to this situation is the question of whether there is a specific special constitutional weakness which we can label *dyslexia* or by some similar term. Neurologists like Orton (34), Hinshelwood (24), and Schilder (36) have attempted to describe and localize the neurological concomitants; Schilder's speculations, in particular, are well worth reading. The persistence of atrocious spelling years after successful response to remedial help in reading and despite continuing special help and good effort, makes it tempting to ascribe at least some cases to a fundamental constitutional special weakness in perceiving word forms and retaining memory images of them. Our belief that there are at least a few cases of constitutional predisposition for reading disability has been reinforced by finding several families in which almost identical learning difficulties are present in closely related individuals.

B. THE DIAGNOSTIC PROGRAM

We find that a comprehensive psychological examination of a reading disability case usually requires five to seven hours. The major components are as follows: (*a*) *Intelligence.* For pre-adolescents, the Revised Stanford-Binet and the *WISC* Performance Scale or Cornell-Coxe Performance Scale. For adolescents, the Wechsler-Bellevue may suffice. We are not yet satisfied with the verbal part of the *WISC* as a substitute for the Stanford-Binet with pre-adolescents. (*b*) *Personality evaluation.* Interview with the child and a Rorschach make a basic minimum; additional projective techniques such as *TAT* (30) or *CAT* (1) drawings, sentence completion, play session. A parent interview is necessary if not performed by social worker. (*c*) *Special Diagnostic tests.* Harris Tests of Lateral Dominance (20), Bender Visual-Motor Gestalt Test (2), Keystone Visual Survey (26). (*d*) *Evaluation of reading skills.* Silent reading tests, oral reading tests, sample reading lessons, or informal testing as needed.

The time required for the different parts of an examination will vary considerably from case to case, and much more can be accomplished during a given time with some children than with others.

We consider it important to include non-verbal as well as verbal tests of intelligence, because reading is a complex skill in which perceptual keenness is significant, and because important clues to causation are sometimes shown by very inferior performances on non-verbal tests. Each of the special tests has a definite purpose in the examination. The lateral dominance tests give evidence of directional confusion in quite a large proportion of reading disability cases. The Bender, while in need of much additional research, shows up marked disturbances of visual perception or visual-motor coördination in some cases, giving important indications of organic disability or of visual-motor immaturity. The Keystone Visual Survey is a series of stereoscopic tests that sometimes disclose visual conditions, not found in the usual school vision tests, that require further medical diagnosis.

The Rorschach is very helpful, both as an indicator of perceptual trends and as a measure of the nature and severity of personality disturbance. The *TAT*, drawings, and other projective techniques are often valuable aids. From material obtained in this way, it is often possible to judge the degree of emotional resistance to learning, and to get into the dynamics of the child's interpersonal relations and attitudes toward himself. Interview techniques are generally similar to those used in other clinical interviews. One should try to make sure that all of the following special points are covered: (*a*) birth and early infancy, with special reference to the possibility of organic brain injury; (*b*) speech development; (*c*) early school experiences and attitudes; (*d*) present school experiences and attitudes; (*e*) attitude toward reading; (*f*) possibility of converted lateral dominance; (*g*) attitude of the family toward the disability; (*h*) amount and kind of help that has been provided by the family and the school; (*i*) presence of similar or related difficulties in other members of the family; and (*j*) recreational and cultural interests and attitudes of the child and family. These are not the only points about which information should be obtained, but rather are points which are specially relevant.

C. SELECTION OF READING TESTS

The selection of reading tests should fit the needs of the case, and therefore no uniform battery of reading tests is desirable. It is recommended, however, that a comparatively small number of tests be used again and again, so as to allow the examiner to develop real familiarity with them and their interpretation.

1. *Children at Low First Grade Level*

A test involving oral reading of connected material should be given first, to determine roughly the level of reading competence and give a preliminary idea of the qualitative pattern. For this, one might use the first paragraph or two of the Gray Standardized Oral Reading Paragraphs (17) or the Oral Reading Test which is the first part of the Gates Reading Diagnostic Tests (12) or the first paragraph or two of the Durrell Analysis of Reading Difficulties (7). The new Gilmore Oral Reading Test (15) looks promising. The norms for all of these tests seem to us to be somewhat unsatisfactory at first grade level, since the Gray norms seem to underestimate, the Gates norms to overestimate reading level, and the Durrell norms are based upon rate rather than accuracy. A practical method of recording is described in Harris (21, p. 180). It is desirable to record the exact nature of each error, both for qualitative analysis and for possible later research. At this level the testing of silent reading is not essential, but to establish a base for measuring later improvement, the Gates Primary Reading Test Type 3 (Paragraph Meaning) (11) may be used.

For those who are relatively unfamiliar with reading diagnosis, it is probably desirable to become acquainted with the techniques by administering a fairly complete battery of tests. The Gates Reading Diagnostic Tests (12), the Monroe Diagnostic Reading Examination (29), and the Durrell Analysis of Reading Difficulties (7) are available. With increased experience, the examiner tends to cut down on the number of sub-tests used and to rely more upon qualitative observations and interpretations than upon test scores.

We have tended in recent years to adopt a fairly uniform selection of sub-tests from the Gates battery, including Phrase Perception; Word Perception both flashed and untimed; Reversible Words; Oral Spelling; Reading Small Letters; Reading Capitals; Giving Letter Sounds; Blending Letter Sounds (Auditory Techniques). Any of these tests should be discontinued or omitted if the examiner believes the child to be unable to make further correct responses, or if it seems to be too frustrating. For the rare child who cannot recognize any words at all, going through a list of tests does not make much sense. Some examiners prefer to use the Monroe Diagnostic Reading Examination, particularly with the almost non-reader. The Word Discrimination Test and the Mirror-Reading Test of the Monroe series are sometimes quite helpful as supplementary tests, even if the rest of this battery is not employed. The Durrell tests seem more effective at levels above first grade.

Certain sub-tests from reading readiness tests may be used to determine whether the child is deficient in specific phases of readiness for reading. We generally select appropriate subtests from the Gates Reading Readiness Tests (13) or the Metropolitan Readiness Tests (23).

We have come to rely increasingly on informal approaches to the appraisal of reading skills at these low levels, and we like to supplement the formal tests with samples of reading from graded readers and with a series of short sample lessons, as described below.

2. *Grades 1 and 2*

For children whose reading ability is middle first grade to high second grade, there are several silent reading tests that may be used fairly appropriately. At first grade level, the Gates Primary Reading Tests Type 3, and Metropolitan Primary I are fairly satisfactory. For those at second grade level, the reading comprehension and vocabulary sections of such widely used achievement tests as the Metropolitan Primary II Battery (22) and the Stanford Primary Achievement Tests (25) may be used. The Metropolitan, in our experience, overestimates when the child guesses freely and underestimates those who read very slowly; the Stanford tends to underestimate the reading ability of dull children. In general, silent reading tests at these low levels are none too trustworthy in the case of severely handicapped readers.

For connected oral reading and for intensive diagnosis of reading skills, the same tests are recommended at this level that have been listed above for low first grade readers, excluding the readiness tests. Sample lessons are important in most of these cases, and informal samples of book reading are also quite helpful. Directions for giving such informal book tests may be found in Harris (21, pp. 132-133), and Betts (3, pp. 456-479).

3. *Grades 3 and 4*

When the introductory oral reading test indicates that reading skills are probably at third or fourth grade level, a somewhat more detailed analysis of silent reading skills is desirable. At these levels we prefer the Metropolitan Elementary (22) or the Stanford Primary Reading Test (25) supplemented by the Gates Basic Reading Tests Type *A*, Reading for General Significance and Type *D*, Reading for Details (10). Other combinations of tests may also be profitably employed. This combination, however, gives useful measures of level of comprehension, meaningful vocabulary, and rate and accuracy in two important kinds of actual reading.

For intensive analysis of oral reading skills the preferred diagnostic tests would be selected from the Gates battery or the Durrell battery. At these levels there should be some inquiry into knowledge of the sounds of syllables and phonogram combinations. If, however, the major difficulties seem to lie in silent reading skills rather than in oral reading and word recognition, testing of oral reading can be reduced to a test of connected oral reading and a test involving recognition of isolated words. Sample word-study lessons are usually unnecessary. Informal tests based on reader and workbook selections are desirable.

4. *Grades 5 and 6*

The attainment of a fifth or sixth grade score in oral reading usually means that the primary difficulty is in silent reading, and therefore a detailed analysis of basic silent reading skills is desirable. However, inability to work out the pronunciation of polysyllabic words and misreading of longer words may sometimes be at the root of what seems to be a silent reading difficulty according to test scores. A limited meaningful vocabulary may also be important in the causation of such a pattern. The problem is how to get a useful and reasonably accurate picture in a relatively short time. Some of the longer and more reliable tests for school use, such as the Iowa Tests of Basic Skills (28), are too time consuming for routine clinical use.

It is important to determine the grade level at which adequate reading comprehension can be maintained; for this, the paragraph meaning section of the Stanford Intermediate Test (25), the Nelson Silent Reading Test (31), or the Metropolitan Intermediate (22) will do. These take about 20-25 minutes. The Gates Basic Reading Tests, Types *A* (6 minutes) and *D* (8 minutes) serve as fairly useful supplements. The first sub-test of the Iowa Elementary Reading Test (18), may also be used as a fairly satisfactory measure of rate and comprehension when supplemented with the paragraph comprehension sub-test from the same booklet.

When an unusually slow rate of reading is discovered, informal observation of eye-movements is desirable. For this, a book should be selected which is about one grade lower than the level indicated by the Stanford or similar tests. Directions for such informal testing may be found in Harris (21, pp. 451-456).

5. *Junior and Senior High School Levels*

For students whose reading is above sixth grade level (and this means reading ability, not grade placement) a survey test such as the Stanford Advanced Reading (25), Metropolitan Advanced Reading (22), or Nelson-Denny (32) would indicate the general level of reading. For more detailed analysis of rate, comprehension, vocabulary, and other silent reading abilities, one of the following is recommended: Iowa Elementary (18), Traxler Silent Reading Tests (37), or Iowa Advanced (19, 27). The results of the survey test would serve as a guide to the selection of the appropriate analytical test. For evaluation of oral reading, one might use the last page of the Gray Oral Paragraphs to determine if there is any deficiency in oral reading skills; if there is, a more detailed analysis of oral reading might be necessary. When slow rate is a major problem, eye movements should be observed as described above. Photography of eye movements during reading is a luxury which may be employed if available, but it is not really necessary.

6. *College Level*

The Iowa Advanced Reading Test (19) combined with the Nelson-Denny Reading Test (32) provides a fairly adequate measure of silent reading skills. The same general directions as those discussed in the preceding section apply here excepting that standards of expectation would necessarily be at a higher level.

Careful analysis of silent reading tests is equally as important as that of oral reading diagnostic tests. Sub-test scores should not be taken too literally without careful scrutiny of the student's performance to determine the real meaning of the scores. For example, rate scores should always be studied along with comprehension results. If the rate of reading sub-test reveals a high score but is accompanied by poor comprehension, the rate score is spuriously high. Likewise, if analysis of the comprehension test reveals that the student showed a high degree of accuracy but was unable to cover much material because of excessively slow rate, the low score should not be interpreted as poor comprehension. In such cases it is usually advisable to give a survey test which has a long time allowance, in addition to an analytical test, in order to obtain a more adequate measure of power of comprehension. Likewise, inferior performance on vocabulary tests should be analyzed to determine whether the poor results are due to inadequate background of word meaning or whether deficiency in word recognition was responsible for the low score. At secondary school level, many students have difficulty in recognizing technical vocabulary relevant to specific subjects and therefore are unable to choose correct meanings on multiple choice tests. If examined orally, these students may be able to define the same words accurately. These errors in interpretation of test results have been found frequently enough to deserve mention.

It is also desirable to subject silent reading test records to some qualitative study. The pattern of right and wrong answers may be quite revealing in regard to degree of consistency of attention, effort, and understanding. Some scores when examined closely turn out to be made up of the accumulation of credits more by guessing than by actual reading. The slow, painstakingly accurate reader is very easily distinguished from the hasty, careless reader. Sometimes it is quite fruitful to ask the child to read orally particular paragraphs on which he has completely missed the point, and in this way with further questioning to get a real understanding of what went wrong in his reading.

D. Informal Lessons as a Diagnostic Technique

An experimental reading lesson is a very helpful part of the diagnostic examination. The clinician can obtain much supplementary information that is valuable both diagnostically and in making recommendations for treatment. This kind of session is very different from the testing situation, which indicates the child's reading status but provides few clues as to his amenability to instruction. The testing situation is a controlled one and is subject to specific directions for administration. Now the child is placed in an actual learning situation. The examiner becomes a teacher and seeks to determine which techniques may be employed successfully in teaching the child to read. The entire session is informal and permissive. There is a considerable amount of give and take and flexibility in approach.

The examiner should note qualitative observations about the child's ability to profit from reading instruction. Such factors as quick or slow grasp, tempo of work, need for repetition of instructions, degree of motivation required, resistance to specific materials, ability to exert effort, to give sustained attention, and to recall what is taught, are of significance in interpreting the results of sample lessons.

Several procedures should be tried until success with at least one method is clearly apparent. If time allows, a variety of materials should be presented so as to obtain some impression of the child's reactions to readers, workbooks, and game-type devices. Any success should be liberally praised and failure minimized. In such a brief period we have been able to demonstrate to many children that they have the ability to learn to read. We have found that, in most cases, this experience has proven to be a powerful motivating force for future reading instruction.

The procedures that are described below are to be used at the discretion of the examiner, following careful analysis of the diagnostic reading test results. Thus, which methods to use and what level at which to begin the sample lessons will be based on an appraisal of the child's basic reading skills as indicated on the tests.

1. *Visual Method*

For the child who is a non-reader or almost a non-reader it is advisable to begin with a simple visual approach to word recognition which is essentially one of learning words by means of picture clues.

For materials, one needs several cards, each with one picture illustrating a well-known object such as a cake, a window, a table, a book, etc. The word is printed or typed under the picture. On another set of cards the words are printed without pictures. Test the child to make sure he does not already know the words, and select about five for teaching.

Present the first picture card, point to the word, and tell the child the word. Ask him to say the word several times while looking at it. Then ask him to find the non-illustrated card with the same word on it. When the child thinks he has learned the word, proceed to the next illustrated word. After five words have been studied, shuffle the five non-illustrated cards and test him on them. If he is successful, re-test about 30 minutes later.

It may be desirable to follow up by teaching a few words in a preprimer workbook, followed by reading the corresponding preprimer. If the child objects to this material, it may be necessary to avoid using readers in the early stages of remedial instruction.

If the child cannot succeed with a visual approach of this sort, it will probably be necessary to use either a visual motor or kinaesthetic procedure in the early phases of remediation. If none of these three is successful, one is probably dealing either with a severe emotional blocking requiring psychotherapy, or with an organic neurological condition similar to an aphasia. In the latter case, progress is likely to be very slow, and work on a reading readiness level may be needed for a long time before the child is capable of learning and remembering word forms.

2. *Word Family Approach*

If the child can grasp a visual procedure readily and shows some knowledge of letter sounds a Word Family method may be tried. This technique of word analysis is especially useful for those children who possess only rudimentary blending ability and who are not yet able to cope with a letter by letter sound-blending procedure described under Phonic Method. It affords a limited degree of independence in word analysis. For example a known word such as "man" may be transformed into "fan, can, ran," by changing the initial consonant. The words learned in the visual lesson may be developed in the same way: "cake, lake, bake, rake, take; book, look, cook, took, hook, etc.

3. *Phonic Method*

If a child shows ability to blend together three sounds that are presented orally such as *b-a-t,* it is safe to try a phonic approach. Teach or review the sounds of about four consonants and the short vowel *a*. The sounds of the letters are presented singly as follows: Show him the letter *m* (use lower case letters). "This is *m* and the sound is mmm as in man. Now I am going to give you some other words that begin with *m*. Listen and try to hear the *m* at the beginning of each word: meat, must, make, milk, etc. Can you hear the *m*? Now give me some words that begin with *m*." Help him with suggestions if necessary. Proceed the same way with *f, s, t, d*.

Teach short vowel sound of "a." "This is *a*, and the sound is ă as in apple."

"Now I am going to say a word quickly, then slowly. Listen. The word is 'cat.' Now I am going to say it slowly 'c-a-t.' This is how it looks in writing." Use anagrams to form the word "c-a-t," "cat." "See if you can put them together to make 'cat.'" Ask him to sound the word out as he does it. Help him, if necessary. Then show him how you can change it to mat, fat, sat, etc.

Write a sentence containing the words, such as "The fat cat sat on the mat."

Change the final sounds: sat to sad, mat to mad, mad to man, fat to fan. Dictate the words to ascertain whether he can write them and sound them as he writes them.

Present words in mixed order: sat, fan, mad, bat, etc. If he is able to read them, there is evidence that he is likely to succeed with a phonic approach.

If a child shows lack of blending ability or is unable to name words that begin with letter sounds it would be advisable to postpone the use of a phonic procedure until after he has had training in auditory discrimination and sound-blending.

4. *Visual-Motor Method*

Choose about three words with which the child is unfamiliar (about 5 to 8 letters in length) such as *friend, airplane, pilot*. Present each word separately. Print the word clearly on a card. Say, "This word is 'friend.' Take a good look at it. What is the word? Now close your eyes. Can you see it with your eyes closed? Look again. What is the word?" Remove the word and ask him to write it. Have him compare his word with that on the card. If incorrect, repeat the above procedure, present the word, pronounce it, have him try to visualize it and then write it again. Sometimes it is necessary to show the card several times before he is able to write it. If he has much difficulty with a word, try another. If he reproduces the word correctly, have him write it again, covering up his previous writing so as to be sure he is recalling the word from memory rather than merely copying it. Check each time to see that it is done correctly. After a period of time has elapsed, review the words.

5. *Kinaesthetic*

The brief summary given below is based on the method fully described by Fernald (8). Only the initial stages used in teaching by a Kinaesthetic procedure are presented here. If the method is adopted, Fernald's book should be consulted.

A short period of orientation is suggested. Tell the child that you are going to teach him to read by means of an entirely new method. Assure him as to its value by telling him that other people who have had difficulty with their reading learned in this way. Describe the procedure to him.

Write the word with a crayon on paper in large size script, or in print (letters approximately 2 inches high). The child traces the word with his index finger (or index finger and thumb, if he wishes) saying each part of the word as he traces it. He repeats this as many times as necessary in order to write the word without looking at the original one. When he appears to know it, he writes it on another sheet of paper. In cases of error or if the child hesitates and seems unable to complete the word, he retraces the word as a whole. He is not permitted to erase in order to correct errors. If he has difficulty recalling the word he should be encouraged to trace it over and over and then to write it without consulting the original writing.

It is evident that any approach may be used as a starting point so as to ensure success in learning to read from the very outset of remedial instruction. However, before long, such procedure must be supplemented by others because a successful reader must have a variety of techniques at his command. It is not necessary to try all five methods listed above. It will be found that most children can learn by at least one of the first three methods. In such cases it is usually unnecessary to try Procedures 4 and 5. However, in cases of severe disability where children cannot learn by a simple visual approach and cannot synthesize sounds together as required in a phonic procedure, the Visual-Motor and Kinaesthetic methods should be tried.

Trial lessons may be used similarly at all levels and should be varied according to the needs of the case. Appraisal of word recognition techniques is usually necessary until the child is functioning around fourth grade level. Since grade scores on oral reading tests are often misleading, informal measures are suggested for determining the appropriate level to begin instruction in oral reading. This can be done by having the child read short selections from about three different levels of a series of graded readers as mentioned earlier.

Trial lessons at intermediate and advanced levels may be used to determine the pupil's grasp of advanced word analysis methods and also his ability to utilize a variety of reading skills. Recognition of technical vocabulary may be tested through a brief sampling of words chosen from published lists containing social studies, geography, history, science, and literature vocabularies (5). Ability to comprehend different types of materials may be ascertained through the use of short selections chosen from workbooks.

The suggestions outlined above point out some of the ways in which an experimental lesson may prove valuable. The examiner will, of course, plan such a session according to the kind of information that he seeks.

E. Recommendations

Recommendations in reading cases generally follow the same principles as apply in other cases examined, except for the addition of specific suggestions for treating the reading problem and for alleviating its attendant complications in the school situation.

Since the etiology of reading disabilities is not always clear, the clinician is often unable to determine the weight to be given to each possible cause. Where any factor appears to have a bearing on the problem, some attempt should be made to treat it. Defects of hearing, vision, or other health problems should be referred for further medical diagnosis and treatment. Where the emotional disturbance is severe enough to warrant psychotherapy, referral for such treatment should be made. In such cases, it is often desirable to ask the therapist to decide when the child may be ready for remedial teaching.

A question is frequently raised as to whether the reading disability was a cause or result of the emotional disturbance. It is generally safe to assume that where there is a history of emotional difficulties since early childhood, the reading retardation is merely one more symptom of the disturbance and treatment of the deeper sources of the problem is indicated. Where the reading disability seems directly related to unsatisfactory methods of handling on the part of the parents or of conflict in the home, referral to a family agency is often advisable. In most cases it is desirable for the parents to discontinue all attempts at teaching the child to read, as we have found that such tutoring usually intensifies rather than alleviates the problem for those children whose parents have been unable to help them before they came to our attention.

The school frequently seeks guidance regarding its rôle in assisting the child to a more satisfactory adjustment in the classroom situation. Suggestions to the teacher should be limited to those that may be carried out without interfering with the regular program of the class. The psychologist, in formulating such recommendations, should be guided by the child's reactions, hobbies, special aptitudes, and specific interests. Suggestions would include emphasis on any positive abilities such as drawing, manual skill, dramatic interests, etc. As for the child who shows no special abilities and also for the one who is restless and hyperactive, minor responsibilities in the class room, monitorial duties, and running errands may provide outlets for excess energy and at the same time give the child a sense of contributing and belonging to the group. In the case of the older pupil who is bright and grasps work easily but whose reading disability prevents him from succeeding on written tests, oral recitation of lessons and, if feasible, oral examinations would bolster his morale considerably. In general, the objectives of recommendations to the school are to encourage an accepting attitude on the part of the teacher towards the child despite his reading retardation, to help alleviate the child's anxiety regarding his school failure, to develop in the youngster feelings of self-worth, and give him status among his peers. If at all possible, a coördinated plan of treatment should be worked out with the school, the psychotherapist, the remedial teacher, and the parents.

Much depends upon the facilities available for special help in reading. In some school systems, there are competent remedial teachers in the school. Sometimes a classroom teacher is interested and willing to try to handle the problem, and with some guidance regarding methods and materials is able to achieve excellent results with the child. The stimulation of being in a group is fairly desirable for some children, while for others, private sessions with a patient and accepting tutor help to relieve the intolerable tension of previous reading experiences and make it possible for an emotionally disturbed child to begin learning.

If a psychologist is to try to give specific suggestions regarding the child's reading program, it is highly desirable that he should make a special study of the materials and procedures that are used in remedial reading. Otherwise, it is much safer for him to point out the child's abilities, present apparent level of functioning, and major needs, and leave the formulation of the actual teaching program to the person who is to carry it out.

References

1. Bellak, L. Children's Apperception Test. New York: C.P.S. Co., Gracie Station.
2. Bender, L. Visual Motor Gestalt Test. New York: Amer. Orthopsychiatric Assn.
3. Betts, E. A. Foundations of Reading Instruction. New York: American Book, 1946.
4. Betts, E. A., & Betts, T. M. Index to Professional Literature on Reading. New York: American Book, 1945.
5. Cole, L. Handbook of Technical Vocabulary. Bloomington, Ill.: Public School Publ., 1940.
6. Cornell, E. L., & Coxe, W. W. Cornell-Coxe Performance Ability Scale. New York: Psychological Corp.
7. Durrell, D. D. Durrell Analysis of Reading Difficulty. Yonkers: World Book.
8. Fernald, G. M. Remedial Techniques in Basic School Subjects. New York: McGraw-Hill, 1943.
9. Ford, F. R. Diseases of the Nervous System in Infancy, Childhood, and Adolescence. Springfield, Ill.: Thomas, 1944.
10. Gates, A. I. Gates Basic Reading Tests. New York: Bureau of Publications, Teach. Coll., Columbia Univ.
11. ———. Gates Primary Reading Tests. New York: Bur. Publ., Teach. Coll., Columbia Univ.
12. ———. Gates Reading Diagnosis Tests. New York: Bur. Publ., Teach. Coll., Columbia Univ.
13. ———. Gates Reading Readiness Tests. New York: Bur. Publ., Teach. Coll., Columbia Univ.
14. Gesell, A., & Amatruda, C. Developmental Diagnosis. New York: Hoeber, 1941.
15. Gilmore, J. V. Gilmore Oral Reading Tests. Yonkers: World Book.
16. Goldstein, K., & Scheerer, M. Goldstein-Scheerer Tests of Abstract and Concrete Thinking. New York: Psychological Corp.
17. Gray, W. S. Gray Standardized Oral Reading Paragraphs. Bloomington, Ill.: Public School Publishing.
18. Greene, H. A., Jorgensen, A. N., & Kelley, V. H. Iowa Silent Reading Tests, Elementary. Yonkers: World Book.
19. ———. Iowa Silent Reading Tests, Advanced. Yonkers: World Book.
20. Harris, A. J. Harris Tests of Lateral Dominance. New York: Psychological Corp.
21. ———. How to Increase Reading Ability (2nd Ed. Rev.). New York: Longmans, Green, 1947.
22. Hildreth, G. H., *et al.* Metropolitan Reading Tests. Yonkers: World Book.
23. Hildreth, G. H., & Griffiths, N. L. Metropolitan Readiness Tests. Yonkers: World Book.
24. Hinshelwood, J. Congenital Word Blindness. London: Lewis, 1917.
25. Kelley, T. L., *et al.* Stanford Reading Tests. Yonkers: World Book.
26. Keystone Visual Survey. Meadville, Pa.: Keystone View.
27. Kohs, S. C. Kohs Block Design Test. Chicago: Stoelting.
28. Lindquist, E. F., *et al.* Iowa Every-Pupil Tests of Basic Skills. New York: Houghton Mifflin.

29. Monroe, M. Monroe Diagnostic Reading Examination. Chicago: Stoelting.
30. Murray, H. A. Thematic Apperception Test. Cambridge, Mass.: Harvard Univ. Press.
31. Nelson, M. J. Nelson Silent Reading Test. New York: Houghton Mifflin.
32. Nelson, M. J., & Denny, E. C. Nelson-Denny Reading Test for Colleges and Senior High Schools. New York: Houghton Mifflin.
33. Olson, W. C. Child Development. Boston: Heath, 1949.
34. Orton, S. T. Reading, Writing, and Speech Problems in Children. New York: Norton, 1937.
35. Robinson, H. M. Why Children Fail in Reading. Chicago: Chicago Univ. Press, 1949.
36. Schilder, P. Congenital alexia and its relation to optic perception. *J. Genet. Psychol.,* 1944, **65**, 67-88.
37. Traxler, A. E. Traxler Silent Reading Tests. Bloomington, Ill.: Public School Publishing.
38. Wechsler, D. Wechsler-Bellevue Intelligence Scale. New York: Psychological Corp.
39. ———. Wechsler Intelligence Scale for Children. New York: Psychological Corp.

Educational Clinic
Queens College
Flushing, New York

Educational Clinic
City College of New York
Convent Avenue & 140th Street
New York City 31

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 341-345.

FAMILIAR SIZE AND SUBCEPTION IN PERCEIVED DEPTH*

Department of Psychology, Cornell University

CAROL BARNES HOCHBERG AND JULIAN E. HOCHBERG

In an earlier paper (3) the writers described an experiment in which *familiar size,* the only depth cue which by definition demands an empiricistic explanation, appeared to be ineffective in influencing reversible perspective when *relative size* was held constant. Moreover, the literature which seemed to support the importance of familiar size in *immediate* perception was found inconclusive by three criteria: either (*a*) familiar size was experimentally confounded with relative size, or *changing size* (neither of which *require* past experience), so that it was unclear which variable was responsible for the results; or (*b*) "demonstrations" of familiar size, while intriguing and cleverly conceived, were lacking in sufficient reference to methods of measurement (and, frequently, the number of subjects, data, and experimental design as well); or (*c*) the procedure required fine discriminations, and invited modification of the initial percept by allowing the subject to search back and forth from stimulus to comparison object, correcting until "satisfied" with his judgment (which may be the resultant of a task set operating on existing perceptual organization, but not necessarily characteristic of the original percept).[1]

The paper was criticized recently by Ittelson (4), who raises several points with regard to our argument, experiment, and conclusions. Primarily, Ittelson seems concerned with our evaluation of the literature. His feeling is that, on the contrary, there is ample evidence for the cue of familiar-size which we seem to ignore or else dismiss peremptorily. Some clarification appears to be in order.

Two of the references he cites (Ittelson, 5; Kilpatrick and Ittelson, 8), were in fact omitted, for they seemed then, as now, demonstrations (see Criterion (*b*) above) which are at present inconclusive. Since they present no statement of the methods of measurement, etc., no comment is possible.

Typical of experiments which Ittelson feels are free of the objections

---

*Received in the Editorial Office on May 31, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] "Where observation is difficult, or when one is forced to consider magnitude as separately measurable entities and not as single integrated experience, *inspection* or close scrutiny tends to occur, using some concept of magnitude which is essentially ideational and cannot be directly perceived" (3, p. 109n; cf. Vernon, 12).

to other studies in this area is one (6) in which a single familiar object ". . . viewed monocularly in an otherwise dark room . . . is perceived, immediately and unequivocally, at some definite distance, which can be correctly predicted on the basis of the familiar size of the object" (4, p. 236). Again, the conclusion that familiar size has operated in immediate perception is by no means clearly justified. This study falls in Category (*c*) above, and lest our criteria for regarding it as inconclusive seem obscure, let us consider the method of measurement and the experimental instructions. The subjects are instructed to match an object in a comparison-field to the appropriate distance of a stimulus in an experimental-field, ". . . using alternate presentation with first one field and then the other viewed . . . approximately 1 sec. with a 1-sec. dark interval between, this sequence continuing *until O was satisfied with his setting*" (6, p. 59; italics ours). The *immediacy* of O's perception is in question here, since there was no immediate comparison field, and since the prolonged judgment period permitted subsequent search, modification, and self-correction. We may even wonder, for that matter, how *perceptual* was this response; the response might perhaps have been made just as well in the absence of the visual stimulus, the subject "imagining" where some familiar object would have to be placed to be appropriate as to distance, etc. (cf. Pratt, 11).

The writers have no intention of setting up hair-splitting definitions of precisely what is to be considered immediate perception as opposed to ideation or secondary processes. Such a sharp distinction would admittedly be arbitrary and probably pointless. Broadly conceived, however, this distinction becomes important when one wishes to know, for example, whether the subject can somehow take into account [by unconscious recognition, "subception" (10), or otherwise] the meaningful aspects of the stimulus *before* he perceives and organizes the stimulus depth-wise. Where "unconscious inference" or its like is invoked to explain the whole of perception, one is almost automatically implicated in a consideration of levels of response. It must be very clear that what looks like a subterranean inference forming O's percept is something separate from, and not merely confused with, the everyday phenomena of judging and weighing alternatives. In other words, we must be reasonably sure that O's response refers to his immediate or primary perception and is not masked or modified by some super-imposed task with which we know he can cope ideationally.

This question is an important one today. Although not inescapably so, proof of the effect of familiar size on immediate experience would provide support for the unconscious recognition or "subception" apparent in the

"New Look" approach to perception and motivation, where it is suggested not only that, at some level of functioning, recognition of the stimulus and of its learned meaning occurs before perceptual organization (cf. Wallach 13), but in addition, that this recognition may sometimes arouse emotional "defenses" which prevent the stimulus from being consciously perceived at all.

Now, subception, like unconscious inference, is probably by its very nature impossible to disprove satisfactorily, but we are not thereby committed to accept lightly such suggestions that the inner life of our subjects is populated by sensitive, mathematically-minded homunculi. Experiments which permit or even encourage subsequent modification of earlier responses before final measurement is made do not constitute convincing evidence of the inferential nature of the first responses. Of the literature cited by Ittelson, Ames (1), Kilpatrick and Ittelson (8), and Lawrence (9) fall under Category *b* above; Ittelson (6), and Ittelson and Ames (7), under *a* and *c*.

Ittelson asks three questions bearing directly on the writers' reported experiments; in the *first* experiment, a panel of a "reversible-screen" drawing bearing a line drawing of a boy failed to be perceived as nearer more of the time than did a panel bearing an equal-sized drawing of a man, while in the *second* experiment, in which the two panels bore different-sized drawings of the same boy, the larger was seen nearer more of the time than the smaller. He asks whether different results would obtain if we had used percentages rather than the differences in the times each panel was seen nearer. Although there are objections to using percentages here, the results in such terms still hold against familiar-size (however, the significances of the differences are altered: in Experiment I, one group shows no appreciable difference, $t = 0.5$, and the other group now favors the *man* somewhat, $t = 1.6$; in Experiment II the larger boy is significantly favored, $t = 6.2$). Ittelson asks also whether *O*s really identified the drawings as man and boy, or whether perhaps the boy was seen as a man in overalls all the while; after the experiment, *O*'s assigned modal ages of 8 and 40 years to man and boy respectively, indicating appropriate identification.

Ittelson raises a further question: how do we know that the same results would have been achieved if other figures—real ones—had been used. To quote from our own recognition of this very issue: ". . . qualifications are necessary as to generalizability to more "life-like" situations, and as to possible optimal differences in *familiar size* of the represented objects, etc. These problems . . . are, however, also faced by the proponents of this cue . . ." (3, p. 113; cf. also Pratt 11, p. 29), and more research is obviously necessary. Certainly we have not *disproved* familiar size as a depth cue, but we have failed to obtain any measurable effects in the present experimental situation, in which relative size was effective.

Finally, Ittelson contends that we have not really separated the two variables of relative-size and familiar-size.

Here, Ittelson's confusion likely stems from his having quoted our final conclusion (deriving from both Experiments I and II) with respect to the first experiment alone. Certainly, with man and boy of equal size, familiar-size is not the only variable present; the size equality might be a cue to equal distance.

However, this cue should either prevent depth reversals (which it did not), or else its effects should be balanced between the two panels, leaving familiar size free to operate; it was for this reason that Experiment II was performed. Experiment I used figures of the same relative size but different familiar size, while Experiment II used figures of the same familiar size but different relative size. "Experimental separation" was achieved in the sense of "untying" or independent variation, rather than the actual "elimination" of relative size—something which (cf. Helson, 2) is probably impossible.

However, *our purpose here was not to "disprove" familiar-size but to indicate the existence of two usually-confounded possible depth-cues,* only one of which—relative size—proved effective here.

SUMMARY

An earlier paper held that evidence supporting familiar-size was usually confounded with relative-size; with the latter held constant, familiar-size was not found effective there. Literature claimed by Ittelson (4) to support familiar size was examined again, his criticisms were considered, but no reason was found to alter the writers' previous position: though familiar size may affect and modify existing perceptual organizations, extension of this claim to include immediate perception (and to infer unconscious recognition or "subception") is still premature.

REFERENCES

1. AMES, A., JR. Some demonstrations concerned with the origin and nature of our sensations (what we experience): A laboratory manual (preliminary draft). Hanover, N. H.: Hanover Institute, 1946 (mimeographed).
2. HELSON, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.,* 1947, **60**, 1-29.
3. HOCHBERG, C. B., & HOCHBERG, J. Familiar size and the perception of depth. *J. of Psychol.,* 1952, **34**, 107-114.
4. ITTELSON, W. H. A note on "Familiar size and the perception of depth." *J. of Psychol.,* 1953, **35**, 235-240.

5. ———. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285-294.
6. ———. Size as a cue to distance: Static localization. *Amer. J. Psychol.*, 1951, **64**, 54-67.
7. ITTELSON, W. H., & AMES, A., JR. Accommodation, convergence, and their relation to apparent distance. *J. of Psychol.*, 1950, **30**, 43-62.
8. KILPATRICK, F. P., & ITTELSON, W. H. Three demonstrations involving the visual perception of movement. *J. Exp. Psychol.*, 1950, **42**, 395-402.
9. LAWRENCE, M. Studies in Human Behavior. Princeton, N. J.: Princeton Univ. Press, 1949.
10. MCCLEARY, R. A., & LAZARUS, R. S. Autonomic discrimination without awareness: an interim report. *J. of Personal.*, 1949, **18**, 171-179.
11. PRATT, C. C. The rôle of past experience in visual perception. *J. of Psychol.*, 1950, **30**, 85-107.
12. VERNON, M. D. Visual Perception. London: Cambridge Univ. Press, 1937.
13. WALLACH, H. Some considerations concerning the relation between perception and cognition. *J. of Personal.*, 1949, **18**, 6-13.

Department of Psychology
Cornell University
Ithaca, New York

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 347-356.

APPEARANCE AND CHOICE STATUS DURING ADOLESCENCE*

Institute of Child Welfare, University of Minnesota

MERRILL ROFF AND DAVID S. BRODY

A. INTRODUCTION

Appearance as appraised by peer group mates seems likely to be an important predictor variable for some personality characteristics, at some age levels, in the multivariate prediction of multiple developmental outcomes which has been discussed by one of the present authors (6, 7). Studies of adjustment, particularly of peer group social adjustment, during adolescence have strongly suggested that appearance is an important item in group settings at this age. Subjective reactions to somatic characteristics and status, as they occurred in the group of youngsters in the California Adolescent Study, have been reviewed in detail by Stolz and Stolz (8). They point out that: "As with other physical characteristics, boys and girls become increasingly aware of individual facial appearance during the cycle of puberty." Both boys and girls of this age period were disturbed by such unwanted attributes as fatness, unattractive features, poor complexion, and lack of height (boys) or excessive height (girls). At a somewhat older age level in college populations, both case histories and experience in individual counseling indicate that these problems remain sources of concern.

This concern suggests that there should be demonstrable relationships of considerable size with some behavioral criteria. However, for some classes of criterion variables this has been found not to be true. It is well established and generally recognized that appearance, appraised either directly or from photographs, is not an effective predictor of intellectual level, knowledge, skills, etc., which can, as a general thing, be adequately appraised by standard tests. In addition, there have been negative findings in relation to some personality traits, although the criterion traits employed have frequently been rather abstract, and have occasionally been not very adequately appraisable in any manner. Three studies from the literature in these areas will be mentioned here.

Cleeton and Knight (3) had a sample of 28 college men and women

---

*Received in the Editorial Office on June 1, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

rated by 20 friends and by 70 strangers on intelligence and on seven personality traits (frankness, originality, judgment, impulsiveness, ability to make friends, will power, and leadership). The correlation between ratings by friends and by strangers on intelligence was .02; for the seven personality traits, the highest correlation was .32. In one of a series of studies of anatomical signs of aptitude described by Hull (5), photographs of a sample consisting of 40 members of the same college sorority were rated by a group of 20 strangers on intelligence, beauty, and neatness. The correlations between their academic grades and these ratings on intelligence, beauty, and neatness were .11, —.16, and —.15, respectively. However, in spite of fairly marked restriction of range, it was found that ratings of beauty by strangers from photographs correlated .61 with ratings by members of the sorority group. Concerning this he commented (5, p. 119): "The high correlation with beauty is to be expected, since the thing itself, except for coloring, is recorded by the camera." Cook, Humphreys and Murphy (4, pp. 667-668) used appearance ratings in an attempt to predict success in pilot training in the Air Force. Ratings based on appearance were made by test proctors who had had several hours to observe the trainees whom they were to rate. Each trainee was rated by three independent observers on a three-step scale; each proctor rated approximately 60 examinees during each session. Inter-rater agreement was well above chance. For a total sample of 2,228 entering pilot training, "the obtained coefficient between the ratings and the criterion was zero." As a result of these and other studies, discussions of interviewing procedures commonly, and very properly, urge caution in making predictions concerning ability or personality characteristics on the basis of visual and other surface impressions obtained during an interview.

The lack of correspondence between concern about appearance and status on these criterion variables is not surprising when it is realized that they are not the characteristics about which concern is felt by those who regard their appearance as inadequate. Boys or girls dissatisfied with their appearance tend to worry about its effect on their social interactions and do not generally expect it to have an adverse effect on their intelligence. It thus seems likely that appearance would be a better predictor for variables having a social choice component than for ability or proficiency criteria. An indication that this may be so is given in a study by Barker (1), who obtained sociometric choices ("the person you would most desire to have for a seat mate during the semester") among a group of 12 university students of both sexes who had not previously known each other. These had been selected from a total class of 28 students. Choices were expressed within this group at the time of the first meeting of the class, and again three months later. A greater than chance agreement was found between initial choices and those made after an acquaintance of three months. Ratings of appearance were found to be related to sociometric status. The girl who was rated the best-looking member of the group received more choices than anyone else on both occasions.

The main problem of the present study is to determine the extent to which peer-group appraisals of appearance are predictive for a criterion which involves social choice, and which is recognized by the persons involved as being of importance to themselves. Although appearance would seem to play a constant part in the informal visual screening of adolescent social life, there are very few occasions throughout this period in which attractiveness of appearance, as such, contributes to a formal social sorting of a population into groups of some permanence. One occasion in which such formal sorting does occur is the selection of pledges by college social fraternity and sorority groups from the group of entering students. It seems likely that the results of this sorting would be to some extent predictable on the basis of appearance alone.

Many other factors, of course, contribute to the final outcome of being chosen, and of choosing or not choosing, to join a particular fraternity or sorority. Among these are family socio-economic status, including such specific items as amount of money available at the time, and direct familial relationships. Personality characteristics and special skills or proficiencies which are not apparent on the surface would also be operative in some cases. If we were attempting to maximize the multivariate prediction of pledging, items such as these would occupy places of some prominence.

Instead, we were interested in the extent to which a univariate prediction of pledging could be made on the basis of appearance alone. The basic question of this study can thus be stated as follows: to what extent can a group of college student observers predict, on the basis of a relatively rapid inspection of appearance, whether or not the individuals of an incoming class will or will not affiliate with a fraternity or sorority group?

B. PROCEDURE

The data were obtained during freshman orientation week at Montana State University. During this period, which preceded both rush week and the beginning of classes, entering freshmen were scheduled in groups for a variety of activities. Mimeographed alphabetical rosters of the entire class were used in this scheduling; thus any specific group at any one activity at any one time would consist of persons whose names fell on certain pages of this total roster. Observers stationed at any one activity would eventually see the entire group, if they were present during the entire period.

A first, and preliminary, set of observations were made in the fall of 1949. Predictions were made for only part of the class; of this part, about half were rated by a team of three observers, while the other half were rated by only two observers. It was found that in the absence of any instructions as to the proportions of positive and negative predictions which should be made, there was wide variation in the percentages of positive predictions from observer to observer. However, the results obtained were definite enough to lead to a repetition of the procedure a year later with a larger number of judges, who made predictions for the entire entering group. Details of procedure will be given for this second, more carefully designed, administration.

In the fall of 1950, ratings were obtained for the entire group of freshmen participating in the orientation period. Predictions were made at the conclusion of sessions at which college aptitude tests were administered. The judges worked in teams of three, or on occasion four. As the examinees left the room, each was asked to hand his test booklet to a judge at one table, his answer sheet to a judge at a second table, and his pencil to a judge at a third table, giving his name each time so that it could be checked on the alphabetical roster. Each judge located the name given and immediately recorded his prediction that the boy or girl would or would not be pledged to a fraternity or sorority. Since it was necessary to keep the line moving, the judges devoted no more time to making a prediction than was required to obtain a name and make an entry. The 13 judges (six boys, seven girls) were juniors or seniors who were assisting in the general orientation program. All were fraternity or sorority members, and were thus also participants in the criterion decisions for their own sex; however, their number in relation to the total organization membership did not seem large enough to produce a criterion contamination problem.

Statistically, each observer can be regarded as a test item with respect to the examinees for whom predictions were being made. Experience in the first administration had indicated that it would be desirable to attempt to influence the item difficulties involved so that there would not be too great a difference between the proportions predicted for the criterion categories and the actual proportions subsequently occurring. It was estimated that about 60 per cent of the girls would be pledged, but that the proportion for boys would only be about one-third. In such a situation, items with difficulties as low as 10 per cent or as high as 90 per cent "pass" would be relatively inefficient. This was a matter of concern since the total number of "items" was at most 13. To exercise some influence on this, instructions to the observers requested that they attempt to make positive predictions for *more* than one-half the girls, but for *less* than one-half the boys. In spite of these instructions, one observer exceeded the others, and the requested proportion of positive predictions, by a margin large enough to lead to the elimination of her predictions from consideration, as being of insufficient item difficulty. For the remaining 12 judges, the percentages of positive predictions for girls ranged from 50 to 72 per cent, and for boys, from 30 to 59 per cent.

The predictions of these 12 judges were pooled to obtain a total score. Separate scores were also summed for masculine and feminine raters, and for two split-half rating groups evenly divided as to sex. The names of the small number of non-white students were not included. These scores were subsequently compared with rosters of pledges as these became available. The results of these comparisons are shown in the following section.

C. RESULTS

1. *Preliminary Study*

In the first study, only five observers were used, and only a part of the class were rated. At this stage no information was given to the judges concerning the number of boys and girls who might be expected to pledge. One team of three made predictions for all those whose names fell in the block from *R* to the end of the alphabet. A second team of two judged examinees whose names fell between *L* and *R*. Each of the two members of this second team made so many positive predictions as to be relatively undiscriminating, and their results will not be further considered.

From the predictions of the team of three observers, a score consisting of the number of positive predictions was summed for each examinee. A comparison of these scores, which range from zero to three, with subsequent outcome (pledged or not pledged), is shown in Table 1.

Since neither the sample size nor the number of judges are as large as might be desired, any statements made about these results must be tentative in nature. However, some conclusions can be indicated and subsequently compared with the findings of the second study described below.

1. The distributions of scores, which would approximate that of a point binomial if there were little or no agreement among raters, tends in each case

TABLE 1
PREDICTIONS OF PLEDGING IN RELATION TO OUTCOME: PRELIMINARY STUDY

	Score*				
	0	1	2	3	Total
		BOYS			
Pledged	1	7	9	12	29
Not pledged	7	19	13	13	52
Total	8	26	22	25	81
		GIRLS			
Pledged	1	2	15	15	33
Not pledged	13	6	4	6	29
Total	14	8	19	21	62

*Number of positive predictions received.

toward rectangularity. This indicates substantial inter-rater agreement, and suggests that the predictions made are concerned with a reliably ratable characteristic.

2. The pooled predictions have considerable validity. Those with higher scores are pledged more frequently. Validity seems somewhat higher for girls than for boys.

2. *Main Study*

The second administration, conducted one year later, differed from the first in three major respects. (*a*) Predictions were made and validated for almost the entire entering class. (*b*) In an attempt to obtain more nearly appropriate proportions of positive and negative predictions, instructions were given to the raters to provide them with some guidance on this point. (*c*) The number of raters was increased, and each rater was asked to make predictions for the entire sample. With these more nearly adequate circumstances, it is possible to check, and in some cases to elaborate on, the conclusions from the preliminary study which were outlined above.

When the 12 observers were split into two sets (three male and three female in each) and scores for each examinee were summed for each set, the same tendency toward rectangularity of the distributions appeared which had been found earlier, and was somewhat more marked for girl examinees. (This effect for all 12 raters combined can be noted in Table 2.) The raw correlations between ratings by these two sets of observers were .81 for girl examinees and .66 for boys. When it is considered that these are, in effect, correlations between two six-item tests with a very short time limit, this level of reliability can be considered high.

A tabulation of the scores obtained by summing the positive predictions

TABLE 2
PREDICTIONS OF PLEDGING IN RELATION TO OUTCOME: MAIN STUDY

	BOYS ($N = 283$)		GIRLS ($N = 205$)	
Score*	Number of cases	Per cent pledged	Number of cases	Per cent pledged
11-12	27	52	54	74
		)58		
9-10	44	61	35	63
7- 8	54	52	30	57
6	29	38	11	64
4- 5	50	40	26	54
2- 3	44	18	26	38
		)14		)24
0- 1	35	9	23	9
Total	283	39	205	55

*Number of positive predictions received.

made for each examinee by all the 12 raters against the criterion (pledged, not pledged) is given in Table 2. It may be noted that about one boy in eight, and about one girl in nine, had scores of zero or one, and of these, only nine per cent were pledged in each group. For boys, the highest percentage of pledging, 61 per cent, occurs in the group with scores of 9 or 10; for girls, the highest percentage, 74 per cent, occurs in the group with scores of 11 or 12.

To obtain more nearly comparable proportions of the two groups than is given by the class intervals shown, upper and lower quarters of the two distributions can be approximated by combining the two upper intervals for boys (71 cases), the two lower intervals for boys (79 cases), the two lower intervals for girls (49 cases) and by taking the one upper interval for girls (54 cases). Among boys, 58 per cent of the upper quarter, but only 14 per cent of the lower quarter, were pledged. Among girls, 74 per cent of the upper quarter, but only 24 per cent of the lower quarter, were pledged. These values correspond well with the proportions shown above in Table 1, and the indication that these predictions have very appreciable validity against this criterion is confirmed.

It is difficult, however, to make an unequivocal interpretation of the relative validity of predictions for boy and girl examinees, due in part to the difference in the proportions pledged. Biserial correlations are not presented here due to uncertainty as to the effect of the marked departure from normality of these distributions on the interpretation of values so obtained. A comparison of the percentages of correct predictions made by each individual rater indicates some superiority for ratings of girls, but the higher inter-rater agreement regarding girls suggests that this difference would shrink as the ratings are aggregated. The safest conclusion seems to be that if any difference is present it is in favor of the girls, but that in any case there is not much difference in validity between the two groups.

If attention is shifted to sex differences between raters, instead of between those being rated, no significant difference in effectiveness is found. The distributions of percentages of correct predictions by individual boys and by individual girls overlap almost completely, for both boy and girl examinees. The preliminary study had not contained enough observers to permit a check of this point.

D. Discussion

The data taken as a whole indicate clearly that rapid evaluations of appearance show a high level of inter-rater agreement and very appreciable validity against a real-life criterion situation involving peer-group choices. Earlier studies have indicated that there are other types of criteria which would not be significantly predictable from ratings of this kind. However, it seems likely that some additional criteria relating to peer-group status would be correlated with the present set of judgments. Brody (2), in a study of a different sample of girls on this same campus, found that reported frequency of dating appeared related to choice status within a girls' residence hall; both these variables would be expected to show some positive relationship to appearance, although not necessarily of the same order as that found in the present study.

Certain aspects of the present criterion may be noted. No information is available concerning the reliability of the pledging process itself. If it were possible to repeat the entire pledging procedure, it seems probable that there would be substantially the same, but a somewhat different, distribution of persons. As is the case in many practical situations, exactly what would happen with such repetition is not known. A second item is that fraternities and sororities, which have been treated here as homogeneous groups, actually differ in status among themselves. These commonly range from those with high prestige value to those with relatively little, and could be so ranked by persons such as the observers of the present study, with considerable reliability. It is possible that such a refinement of the criterion would sharpen predictability somewhat, but this was not attempted.

Since our main interest lay in the validation of appearance ratings rather than in the multivariate prediction of pledging, no systematic study was made of the contribution of other factors to the results found. Informal coun-selling experience with some members of the sample suggests that undesirable personality characteristics, which were not apparent on the surface, but which may have been known to acquaintances, may have played a part in some cases whose appearance was considered good. Financial status at the time was a potentially important variable concerning which no information was obtained. Specific family relationships ("heritages") may have helped to account for the pledging of some who were low in appearance. Predictions based on a number of these variables would, of course, be expected to exceed that made on the basis of appearance alone. Conversely, to the extent to which these variables were operative, and particularly if they were relatively uncorrelated with appearance, they would tend to lower the predictive value of appearance alone (as, persons adequate in looks but inadequate in funds). However, the univariate predictability which has been found indicates that the concern of adolescents about the social effects of appearance is realistically oriented toward the objective social events of their groups.

E. SUMMARY

The purpose of this study was to determine the extent to which social sorting among adolescents in a real-life situation involving peer-group choices is predictable on the basis of appearance. Comparisons have been made of predictions based on appearance with action subsequently taken in pledging or not pledging to college fraternity or sorority groups. Undergraduate junior or senior students made predictions concerning entering students during the orientation period in two successive years. At the time of the second, and main, study, predictions were made for almost the entire entering class. Predictions were summed for each examinee, and the resulting scores were compared with pledge lists when these became available. Results which had been indicated in an initial study were substantiated and elaborated somewhat in the second. The following conclusions seem indicated.

The reliability of these predictions, as estimated by the amount of agreement between one half of the raters and the other half, was relatively high. In both sets of observations, inter-rater agreement was greater for girl examinees than for boys.

For the 283 boys and 205 girls included in the main study, scores from combined predictions by 12 observers were compared with the criterion (pledged, not pledged). Among boys, 58 per cent of the highest quarter, but only 14 per cent of the lowest quarter, were pledged. Among girls, 74 per cent of the highest quarter, but only 24 per cent of the lowest quarter, were pledged. Generally comparable values were obtained in observations made in the preceding year. There seems to be little difference in the validities found for boys and for girls. Masculine and feminine raters were about equally effective in predicting pledging for both boy and girl examinees.

REFERENCES

1. BARKER, R. G. The social interrelations of strangers and acquaintances. *Sociometry,* 1942, **7**, 169-179.
2. BRODY, D. S. Developmental factors affecting sociality traits and work habits among college women. Unpublished Ph.D. thesis, Univ. Minnesota, 1952.
3. CLEETON, G. U., & KNIGHT, F. B. Validity of character judgments based on external criteria. *J. Appl. Psychol.,* 1924, **8**, 215-229.
4. COOK, S. W., HUMPHREYS, L. G., & MURPHY, R. E. In Guilford, J. P. (*Ed.*), *Printed Classification Tests.* AAF Psychology Program Research Reports, No. 5. Washington: U. S. Government Printing Office, 1947.
5. HULL, C. L. Aptitude Testing. Yonkers: World Book, 1928.
6. ROFF, M. A factorial study of the Fels Parent Behavior Scales. *Child Devel.,* 1949, **20**, 29-45.
7. ———. Intra-family resemblances in personality characteristics. *J. of Psychol.,* 1950, **30**, 199-227.
8. STOLZ, H. R., & STOLZ, L. M. Adolescent problems related to somatic variations. *Yearb. Nat. Soc. Stud. Educ.,* 1944, **43**, 80-99.

Institute of Child Welfare
University of Minnesota
Minneapolis 14, Minnesota

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 357-362.

THE POSSIBLE EFFECTS OF PROJECTIVE TESTING UPON OVERT BEHAVIOR*

Psychiatric Institute, New York City

EMANUEL F. HAMMER

A. THE PROBLEM

Recently the following theoretical questions have been raised by a number of clinicians: Can projective testing be expected, upon occasion, to precipitate an acting out of conflicts? Does projective testing tend to lower, at least momentarily, the repressive or suppressive defenses erected by the subject to control the expression of forbidden behavior? Does the period in which these defenses are lowered extend beyond the testing session? Can the consistency of projective data in highlighting certain conflicts or stress constellations be taken as a clue that the material in question is close to consciousness? And if this last is so, does it tend to prognosticate a proneness toward the acting out of behavior related to the stress producing feelings, in the near future? Can a supportive psychotherapy session, following the projective testing period, operate to avert the acting out of culturally condemned behavior, where projective techniques catalyzed greater awareness in the subject of his anxiety-producing dynamisms?

These questions appeared to have little more than academic interest until recently when a 28-year-old male subject was seen by the writer. The richness and consistency of the material produced by projective testing, and the subject's dramatic behavior immediately following the examination period, served to bring these questions into sharper focus and to suggest certain possible hypotheses as answers to them.

B. PERSONAL HISTORY

The patient was born in a small town in a Mid-Atlantic State in 1923. He was the youngest of four children. Little more than the following could be discovered concerning his early background. He finished four years of elementary school, and was there described as, "a nervous, sickly, underweight child, given to temper tantrums." After finishing school he held various jobs but never worked steadily, and was never satisfied with his job.

---

*Received in the Editorial Office on June 2, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

Six years ago the patient received a head injury when he was thrown through the roof of a car during an accident. No fractures or neurological evidences of brain damage, however, were found. He began to suffer from headaches and about two months later had his first seizure, which was described as being of the typically grand mal type. Neither tongue-biting, incontinence, nor injury to himself, however, has been observed during a seizure. In spite of the seizures, the patient enlisted in the army in 1945 and remained in service for two years. He was successful in concealing his seizures for this extended period. When the epileptic condition was discovered he was discharged, and he then worked intermittently at odd jobs.

On one occasion he was arrested on a charge of attempted rape, and after three months in jail he was transferred to the hospital because of his epileptic condition.

Shortly before the car accident the patient, while on a neighbor's property, was threatened that he would be shot if he didn't leave. The patient told the neighbor that he didn't believe him, and then bent down offering his buttocks as a target. The neighbor fired, and the patient reports: "The bullet went through my rectum and came out through my penis." The content of this statement, in spite of its delusional flavor, turned out to be essentially true. The medical report reveals that there is a healed bullet hole on the inside part of one of the buttocks as well as injured testes and bladder. The penis itself, however, has not been affected. The patient reports that occasionally urination and sexual intercourse are so painful that he has to "scream and stop."

C. Projective Testing

His somewhat contaminated concept of an injured bladder and penis is clearly reflected in his response to Rorschach card No. 10: "It looks like a tree with water underneath in a hole in the ground; the water hole is under the roots and insects are swimming in the water."

Although strong feelings of castration showed up rather clearly on all the projective techniques, they were most markedly evidenced on the *H-T-P*. In his chromatic drawing of the House, the chimney was depicted in distinctly phallic style with the head of the phallus displaced to the right of the chimney proper as if ready to fall off. The chromatic drawing of the Person depicted a man wearing a hat which was also displaced off the vertical axis. His story to *TAT* card No. 17BM involved a man climbing down a rope which suddenly broke. On Rorschach Card III he saw the upper red area as "a snake ready to fall." His wish for, and feeling of

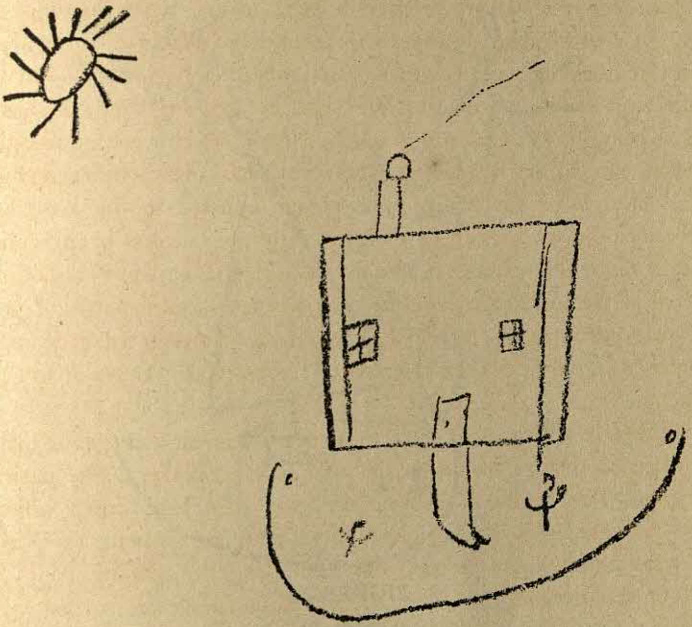


FIGURE 1

House

(House in red, sun in yellow, foreground in black.)

futility concerning, a more adequate phallus was reflected in his story to *TAT* Card 14 (silhouette of a person in front of a window) in which a man stands wistfully at the window watching an airplane "fly by."

The picture presented by the integration of the projective data suggests that the injury to his genitals found fertile ground in an unresolved Oedipal constellation so as to result in the castration feelings so prevalently revealed above. The patient's father died when the patient was 14, a time when the patient's erotic reawakening was presumably at its height. Guilt and fear of retribution were caused by the unconscious feeling that the omnipotence of his hostile wishes toward his father had resulted in the latter's death. The patient has dreams of his mother which are never finished; evidently the superego steps in and interrupts the dream before any forbidden action can take place. The injury to his genitals inflicted by the elderly neighbor was susceptible to unconscious interpretation as retribution for erotic desires toward his mother and death wishes toward his father—a retribution consisting of castration at the hands of an elderly father-figure.

Feminine identification combines with castration feelings to cause this per-

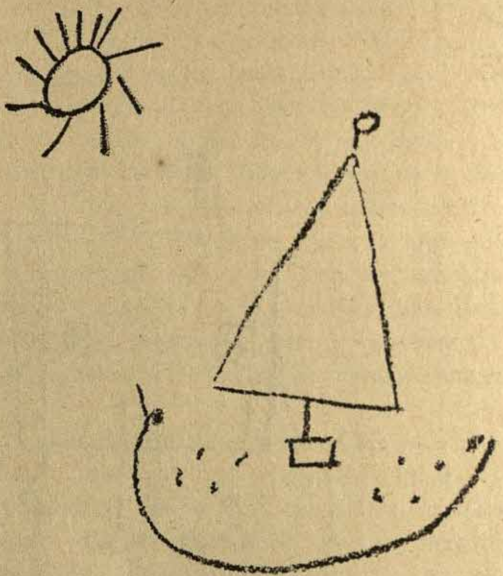


FIGURE 2

TREE

(Tree in red, sun in yellow, foreground in black.)

son seriously to contemplate giving up of the masculine rôle. Contemporary females are viewed with hostility. His repressed tendencies toward homosexuality produce the paranoid symptom consisting of the feeling that people stare at him: "Some people look at your personality—to see if you're clean." One cannot help hypothesizing that this man's homosexual tendencies were unconsciously operative during the shooting incident previously referred to. When the neighbor lifted his rifle and threatened to shoot, the patient responded to the phallic symbol in the neighbor's hands by offering his buttocks—a symbolic homosexual gesture.

Two other bits of behavior fit into the picture. One, the patient's attempted rape may be viewed as a last ditch attempt to convince himself and others of his masculinity. As his object he chose a young girl, a less challenging female whose lack of experience would operate to keep his sexual inadequacies from becoming evident, a girl whose naivete would lead her to be impressed by the masculinity of his attack. Two, some time ago when the patient was unable to find a job—another situation interpreted as a reflection on his masculinity—he contemplated suicide.

Structurally this man's personality is constricted and dependent. In his

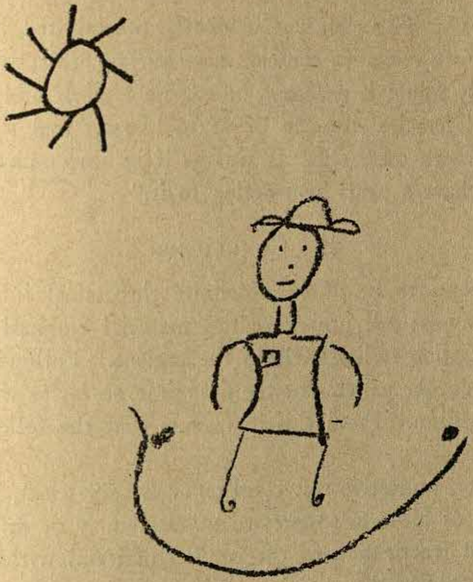


FIGURE 3

PERSON

(Person in black, sun in yellow, foreground in black.)

choice of a Christmas tree when asked to "draw a tree" he indicated that he has regressive tendencies and longs for childhood again; that he desires to return to a period when everything was given to him and he was called upon to give nothing in return other than the privilege of allowing himself to be admired by a doting, worshipping mother. The Christmas tree further reflects his exhibitionistic tendency.

He is presently handling his personality problem by withdrawing into a schizoid, possibly preschizophrenic, state. He has made his personality inaccessible to others and lacks adequate contact with reality. That part of his relationships with people which remains is fundamentally barren and cold. ("Snow" and "ice" are projected to areas on the Rorschach where humans are popularly seen.) Possible delusional material is hinted at in his protocols.

D. POST-TESTING BEHAVIOR

Of unfortunate consequence to the patient, although of theoretical interest to the psychologist, was the confirmation of the patient's castration feelings found on the behavioral level. After leaving the psychologist's office, the patient stopped on the road leading to his ward and exposed himself to a group of females.[1] This act was evidently precipitated by his castration feelings having risen closer to consciousness during projective testing.[2] The patient apparently found it necessary to expose himself and elicit a reaction from a group of females so as to prove to himself and to others that his genitals had not been castrated. It was as if he unconsciously said: "Convince me that I have a penis by reacting to it."

E. CONCLUSION

This case appears to be of very definite theoretical interest because of the following reasons: All projective test material emphasized the presence of profound castration feelings within the patient. Following the intensive examination session, the patient for the first time, so far as we know, resorted to frank exhibitionism. From this it appears that the following hypotheses are tenable.

1. Where inter-projective test consistency is very great, the likelihood of frank acting out of hitherto suppressed material may be anticipated because of the heightened reactivation of the conflict material within the patient.
2. In such instances the clinician's job—even when he is functioning as a diagnostician only—is not terminated with the recording of the last response. He will probably be wise to conclude the interview with a supportive psychotherapy session in which the patient might achieve emotional release and reassurance before leaving the sheltered atmosphere of the examiner's office. This might well prevent the acting out of conflicts which, in turn, might lead the patient into difficulty with the law, the community, or his family, as well as his own conscience.

Psychiatric Institute
722 West 168th Street
New York City 32

---

[1] At an institution for mental defective and epileptic patients.
[2] No interpretations or even reflection of feelings had been given the patient.

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 363-378.

RESPONSE SIMILARITY-DISSIMILARITY AND DIFFERENTIAL MOTOR TRANSFER EFFECT*[1]

USAF School of Aviation Medicine, Randolph Field, Texas

GEORGE T. HAUTY[2]

A. INTRODUCTION

The major objectives of this investigation were to determine for a given psychomotor apparatus: (*a*) The relative difficulty of methods of operation which differ primarily in the axial relationships between stimulus and response. Essentially, the question is: Which is the more difficult, the method or task where the spatial axes of stimulus and response are in common, or where these spatial axes are not in common but orthogonal? (*b*) Differences in proactive facilitation and/or interference occurring in two different combinations of initial-transfer tasks. One such initial-transfer situation is obtained by completely reversing the stimulus-response relationships existing in the initial task so that the initial-transfer response members are directly opposable in space. The other initial-transfer situation does not involve such a reversal. Instead the initial-transfer response members are nonopposable in that they are made on othogonal axes. In all cases the stimuli in the initial and transfer tasks are physically identical.

Of particular importance is the latter objective since its realization constitutes a critical test of certain hypotheses of negative transfer. Recently, it has been contended by Gagné, Baker, and Foster (1) that the negative sign of transfer is determined not by variation of similarity in the initial and subsequent stimuli and responses but by the reversal of the stimulus-response relationships, existent in the initial learning situation. It is this condition of reversal where, if the stimuli and responses of the second task are identical with those of the first task, the strength of the now correct response tendencies is maximally opposed by the now incorrect response tendencies and, accordingly, maximal negative transfer will occur. This hypothesis, there-

---

*Accepted for publication by Ernest R. Hilgard of the Editorial Board, received in the Editorial Office on June 12, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] Reported in greater detail and with other considerations in USAF School of Aviation Medicine Project No. 21-0202-0005, Report No. 2, Jan., 1953.

[2] The writer is indebted to Dr. George Gentry of the University of Texas, Dr. R. B. Payne and Dr. P. E. Fields for their reading and helpful criticisms of this paper.

fore, would predict negative transfer to occur in the case of the opposable initial-transfer situation.

However, if the multidimensional nature of similarity is taken into consideration, as has been done earlier (9, 10, 11), the following assumptions of similarity-dissimilarity may be considered. On the apparatus used, the opposable initial-transfer responses are rectilinearly opposable; hence, with a common spatial axis, there is an aspect of similarity between the initial-transfer response members. The nonopposable situation involves initial-transfer responses which are made on orthogonal axes of the spatial field. Here, since initial-transfer responses are not accomplished on a common axis but on two different axes, dissimilarity may be said to exist.

On the basis of these assumptions of similarity-dissimilarity, a test can also be made of the Wylie hypothesis of negative transfer (13) which is dependent upon the condition of interference given by similarity and dissimilarity of stimulus and response respectively. In direct contrast to the hypothesis proposed by Gagné *et al.,* the Wylie hypothesis would predict maximal negative transfer to occur in the case of the nonopposable initial-transfer situation.

The effect of attentiveness was also to be investigated. Everyday observation reveals the fact that "mistakes" occur most frequently when motor performance is less demanding of directed attention. Assuming substantial relationship between degree of attentiveness and the learning of a motor task, particularly the initial stages, attentiveness should be reduced if less effort or concentration is required in learning. To attempt to determine in the case of a relatively simple motor discriminative task what relationship exists between attentiveness and effectiveness of interference in the transfer task, two sets of instruction were given. In one all the correct responses were indicated for the subject before the initial and transfer tasks and for the other, the correct responses were not indicated.

B. APPARATUS

The psychomotor apparatus used was a slightly modified model of the *SAM* Self-Pacing Discrimination Reaction Time Test (CP611E), a variation of an earlier model which has been completely described (6). A brief description of the apparatus is as follows: On a vertical panel appear four lamps which, if connected by straight lines, would occupy the points of the right angles formed by a square. The upper left and lower right are red and the remaining two are green. A stimulus consists of a pair of lighted red and green lamps of which there are four such combinations.

As can be realized, a particular stimulus is directional in nature either right, left, up, or down depending on whether the red or the green lamp is regarded by the subjects as the cue or referent. Response to the stimulus consists of throwing one of the four spring-loaded toggle switches whose positions on a horizontal panel represent the four points of the compass. Throwing the correct switch results in the extinguishing of the particular stimulus lights and an immediate presentation of a different stimulus, whereas an incorrect selection will result in the particular stimulus lights being extinguished and remaining so for a delay period of two seconds before reappearing to elicit a second response. The cycling mechanism of the control unit presents randomly the four different stimuli for four 2-minute trials which are separated by 30-second periods. Since the correct response determines the presentation of stimuli, the subject sets his own pace.

The modification, earlier mentioned, consisted of a switching circuit which allowed any one of the four switches to be the correct response for each of the four stimuli; thus, four different learning tasks were made possible. Using, for example, the two upper lamps, red left and green right, as the stimulus and considering the red lamp as the directional cue, the correct switch for Task I is in the direction indicated by this cue, left switch; for Task II at a 180° offset or completely reversed in direction, right switch; for Task III at a 90° counter-clockwise offset, bottom switch; for Task IV at a 90° clockwise offset, top switch.

Scoring, correct and incorrect responses, was accomplished by electric counters.

C. PROCEDURE

The experimental conditions are given in Table 1 where it will be noted that Groups *A, D, E,* and *H* learn initial and transfer tasks which involve opposable response members while Groups *B, C, F,* and *G* learn initial and transfer tasks involving nonopposable response members.

Due to the nature of the availability and daily obtainment of subjects and the fact that the groups were not equated, a prearranged scheduling was resorted to so as to permit a near equal distribution of subjects in all conditions throughout the course of the investigation. In addition to the general instructions concerning the operation of the apparatus, the "instructed" subjects were shown all the different stimuli and the corresponding correct switches, e.g., in Task I, "When the red light is to the left of green (the exact lamps being pointed out), push the left switch" (and the left switch was pushed), etc. This procedure for these subjects

TABLE 1
EXPERIMENTAL DESIGN

		Instructed						Uninstructed			
Groups	N	Initial Tasks	Trials	Transfer Tasks	Trials	Groups	N	Initial Tasks	Trials	Transfer Tasks	Trials
A°	16	I	8-2 min.	II	8-2 min.	E°	16	I	8-2 min.	II	8-2 min.
B^n	16	II	8-2 min.	III	8-2 min.	F^n	16	II	8-2 min.	III	8-2 min.
C^n	16	III	8-2 min.	II	8-2 min.	G^n	16	III	8-2 min.	II	8-2 min.
D°	16	IV	8-2 min.	III	8-2 min.	H°	16	IV	8-2 min.	III	8-2 min.

°Perform on opposable tasks. ^{n}Perform on nonopposable tasks.

was also followed before the transfer task, e.g., ". . . now the correct switch you have just learned will not be correct for this phase of the test. The correct switches to be learned are as follows: If the red light is to the left of green, push the right switch, etc." (Task II). The "uninstructed" subjects received the general operational instructions and, in addition, were told before the initial tasks that for each different pair of red and green lights that appeared they were to learn the corresponding correct switch. Before beginning the transfer tasks, these subjects were informed that the correct responses they had just learned would now be incorrect, and that their task now was to relearn the test. All groups were given eight 2-minute trials on their respective initial tasks, with 30-second rest periods between all trials except between the fourth and fifth trials where a 3-minute rest period was given. At the end of the eighth trial, a 10-minute rest was given during which biographical data were taken from the subjects and they were required to "walk around." The apparatus was then switched to the assigned transfer task and the subjects began the transfer task in which the number of trials and rest periods were identical to those of the initial task.

The 128 subjects, obtained from Lackland Air Force Base were male basic trainees, right-handed and high school graduates. Their range of ages was 18-23 years.

D. RESULTS

The initial and transfer curves of correct and incorrect responses for the "instructed" groups (*A*, *B*, *C*, and *D*) are plotted in Figure 1. In the initial task section the four uppermost curves are based on mean correct responses per two minutes and are indicated by type of task. The four curves based on mean errors per two minutes occupy the lower portion of the graph and are also identified by type of task. The curves for mean correct responses exhibit a marked rank ordering with no crossover. In the error curves the only possible difference is occasioned by the departure of the curve for Task IV from the others.

Turning to the section on transfer tasks, the four uppermost curves which are based on mean correct responses per two minutes may be identified by using as an illustration the uppermost curve. Numeral IV at the very beginning of the curve indicates the initial task to have been Task IV and the solid line of the curve indicates the transfer task to be Task III. The four lower curves, based on mean errors per two minutes, are identified by the numerals, e.g., II-III indicates the transfer task to be Task III

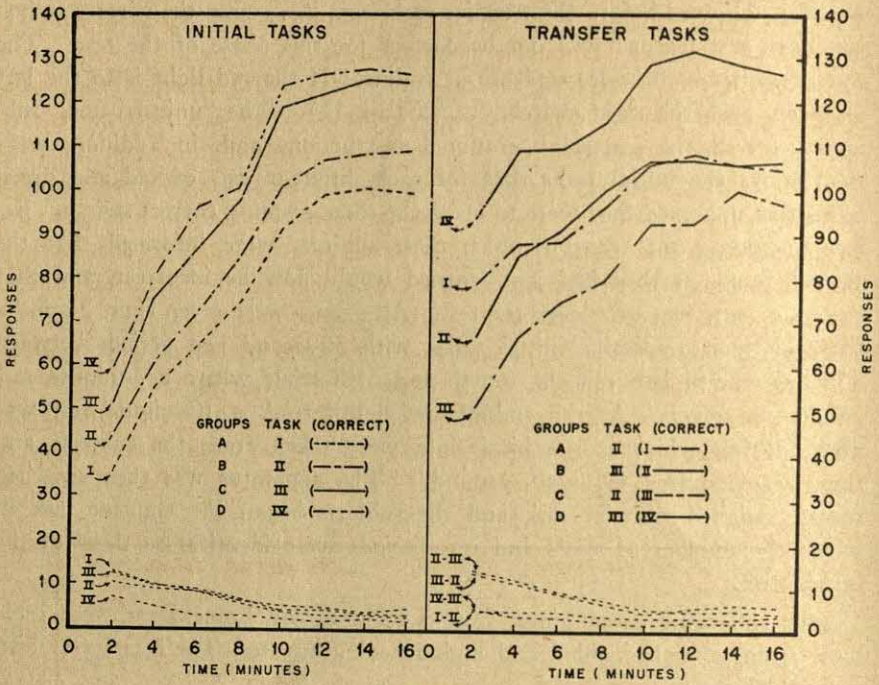


FIGURE 1
Mean Correct and Incorrect Responses Made by the "Instructed" Groups in Initial and Transfer Learning

preceded by initial Task II. Inspection of the correct response curves reveals substantial separation to be a continuous characteristic of the curves of Groups *B* and *D* (transfer Task III) and of the curves of Groups *A* and *C* (transfer Task II). Before this differential transfer effect can be attributed to the similarity and dissimilarity given by the opposable-non-opposable response members, the effects of task difficulty and degree of learning must be considered. Examining first the initial and transfer curves of Groups *B* and *D*, one could not attribute the difference in the transfer tasks to the variable of similarity-dissimilarity in view of the differences existing between the initial tasks. However, in examining the initial and transfer curves of Groups *A* and *C*, one sees, as in the case of *B* and *D*, that the greatest facilitation occurs when the transfer task follows an initial opposable task in spite of the fact that for *A* and *C* the ordering of initial task difficulty and degree of learning is the reverse of that for Groups *B* and *D*. This pattern of results will also be noted for the "uninstructed" groups. The curves based on mean errors likewise exhibit agreement with these findings. The two uppermost error curves are for transfer Tasks II and III which were preceded by nonopposable initial tasks.

Comparing the curves in the transfer section with those in the initial section certain possible differences are indicated: (*a*) All four correct response transfer curves exhibit positive transfer initially, i.e., on the first trial of transfer learning. (*b*) This positive transfer is maintained throughout the remainder of the trials in transfer tasks which have been preceded by initial opposable tasks. (*c*) This initial positive transfer does not appear to be maintained throughout the remainder of the trials of transfer tasks which have been preceded by initial nonopposable tasks.

Figure 2 presents the initial and transfer curves of correct and incorrect responses for the "uninstructed" groups, *E, F, G,* and *H.* The system of identification, type of line, and numerals is that used in Figure 1. In the initial task section appears the rank ordering of tasks demonstrated in Figure 1 with the exception of the changed positions of Tasks III and IV.

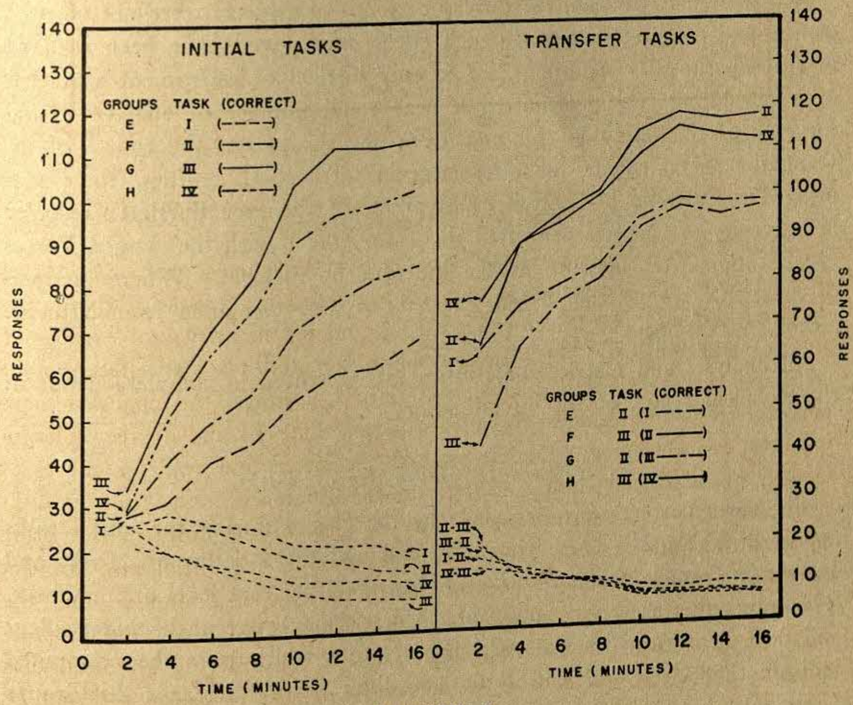


FIGURE 2
MEAN CORRECT AND INCORRECT RESPONSES MADE BY THE "UNINSTRUCTED" GROUPS IN INITIAL AND TRANSFER LEARNING

The differences between the error curves for Tasks I and II and for Tasks III and IV are in accord with those differences noted in the curves based on correct responses. In the transfer task section there is revealed by the correct response curves of the two transfer Type III tasks, Groups *F* and *H*, a small initial difference which graphically becomes nonexistent on the second trial. At this point, in fact, the two curves crossover. A like phenomenon is to be noted in the curves for the two transfer Type II tasks, Groups *E* and *G*, except that here there is no crossover but from the eighth trial on the distance separating the two curves is quite small. The curves based on errors in the transfer section exhibit on the first trial the relationships found in the initial portions of the curves based on correct responses. It is apparent when the transfer Tasks II and III are compared with initial Tasks II and III that positive transfer occurs initially and throughout all four transfer curves.

1. *Statistical Analyses*

As has been mentioned, the different groups of subjects were not equated on the basis of pretesting since such effects might well have been realized in the experimental conditions. The only method of assignment was that of a scheduling which permitted a near equal number of subjects tested for each of the conditions for any given day throughout the course of investigation. To equate for differences in intellectual abilities, there was utilized the method of analysis of covariance (whenever Bartlett's test for homogeneity of variance permitted this assumption) with the Armed Forces Qualification Test measure as the independent variable. Where analyses of covariance were accomplished, the *t*-ratios were computed from adjusted means and variances.

All analyses and matrices of *t*-ratios are presented in complete detail in the technical report of this investigation.

2. *Differential Difficulty of Tasks*

The mean correct responses made by the groups during the initial tasks appear in Table 2. These were analyzed and the *F*-ratios for Tasks and Instructions were significant at better than the 1 per cent level of confidence. The interaction was not significant. Correlation between the independent and dependent variables was .42. The *t*-ratios which were then computed indicate that Groups *A* and *B* do not differ significantly nor do Groups *C* and *D*. Group *D*, however, does differ from *A* and *B* at or better than the 1 and 5 per cent level of confidence respectively. The significance of

TABLE 2
MEANS AND STANDARD DEVIATIONS OF CORRECT AND INCORRECT RESPONSES

	Initial Tasks				Transfer Tasks			
Instructed	A (*I*-II)	B (*II*-III)	C (*III*-II)	D (*IV*-III)	A (I-*II*)	B (II-*III*)	C (III-*II*)	D (IV-*III*)
Correct:								
Mean	77.17	85.59	100.08	104.84	97.24	101.33	82.07	118.04
SD	20.25	29.70	21.48	24.02	23.95	21.91	25.51	20.27
Incorrect:								
Mean	6.68	7.00	7.11	3.68	4.30	8.27	8.21	3.28
SD	5.63	3.77	5.93	2.46	4.10	3.25	4.37	1.82
Uninstructed	E (*I*-II)	F (*II*-III)	G (*III*-II)	H (*IV*-III)	E (I-*II*)	F (II-*III*)	G (III-*II*)	H (IV-*III*)
Correct:								
Mean	48.44	61.01	85.04	75.65	86.56	102.32	80.59	100.98
SD	18.88	25.61	23.86	32.82	24.03	22.84	21.90	33.61
Incorrect:								
Mean	23.13	20.29	13.85	15.31	10.37	10.56	10.59	10.98
SD	9.94	10.94	13.42	12.33	10.05	9.27	8.33	13.19

differences given by Group *C* vs. *A* and *B* does not reach the 5 per cent level of confidence. For the "uninstructed" groups, all comparisons except that of *G* vs. *H*, result in differences significant at or better than the 5 per cent level of confidence.

The consistent variance difference between the incorrect responses made by the "instructed" and "uninstructed" groups was of such magnitude that a test of homogeneity ($P = .01$) did not permit the assumption of homogeneity of variance. Accordingly, the instructed and uninstructed groups were analyzed singly. The resulting *F*-ratios did not meet the 5 per cent level of confidence.

3. *Effects of Opposable and Nonopposable Initial Tasks on the Transfer Tasks*

The mean correct and incorrect responses made in the transfer tasks are also given in Table 2. The mean correct responses were analyzed and the *F*-ratios for Tasks and Instructions yielded probability values of $< .01$. The interaction was not significant. Correlation between the independent and dependent variables was .45. The *t*-ratios which were computed indicate for the "instructed" groups that the differences between *A* and *C* (transfer Task II) and between *B* and *D* (transfer Task III) were significant at or better than the 5 per cent level of confidence. In both cases, the superior group was that which had learned initially a task which was opposable to the task learned in the transfer situation. Differences between the "uninstructed" Groups *E* and *G* (transfer Task II) and between *F* and *H* (transfer Task III) were not significant.

If, for both "instructed" and "uninstructed" groups, a comparison is made of Task II vs. Task III when the preceding initial tasks are opposable, significant differences are obtained. Likewise, when the preceding initial tasks are nonopposable, significant differences occur.

Because of the heterogeneity of variance, the errors made by the "instructed" and "uninstructed" groups during the transfer task were analyzed singly. The *F*-ratio for the "uninstructed" groups could not be considered as being significant. For the "instructed" groups, an *F*-ratio significant at better than the 1 per cent level of confidence was obtained from an analysis of variance. The *t*-ratios indicated that the errors made by Groups *A* and *D* who learned opposable initial tasks were significantly less than those made by *B* and *C*.

For the "uninstructed" groups no significant differences were found for opposable vs. nonopposable initial task effects when the group mean performances for all the trials were tested. To determine if significant differences do exist during the initial portion of the transfer learning curves, the mean correct responses made on the first transfer trial were analyzed. For Groups *E* and *G*, which learned transfer Task II, significantly more correct responses were made by *E* which had previously learned an opposable initial task. The significance of difference between *F* and *H*, however, did not reach the 5 per cent level of confidence.

4. *Test for Transfer*

Task III was learned by Groups *C* and *G* as the initial task, by Groups *B* and *F* who had first learned initial nonopposable Task II, and by Groups *D* and *H* who had first learned initial opposable Task IV. Accordingly, the mean correct responses made during the initial task by Groups *C* and *G* and during the transfer tasks by Groups *B, F, D,* and *H* were analyzed. The *F*-ratios for Groups and Instructions were significant at or better than the 1 per cent level of confidence. The correlation between independent and dependent variables was .41. The *t*-ratios revealed that *C* and *B* did not differ significantly but *C* and *D* did differ at the 5 per cent level of confidence which indicates positive transfer to have resulted in the learning of Task III following the learning of the initial opposable Task IV. While the significance of difference given by Groups *G* vs. *F* and *G* vs. *H* did not reach the 5 per cent level of confidence, the probability values were 0.11 and 0.07 respectively.

Groups *B* and *F* learned Task II initially. Groups *A* and *E* learned Task II after having learned initial opposable Task I and Groups *C* and *G* learned Task II after having learned initial nonopposable Task III. An analysis of covariance of the correct responses made by these groups yielded a significant interaction which did not permit significant *F*-ratios for tasks or instruction so the "instructed" and "uninstructed" groups were analyzed singly. The *F*-ratios were significant at the 5 per cent level of confidence and t-ratios were then computed. The difference between Groups *B* and *A* was significant at a 0.06 level of confidence. Groups *B* vs. *C* did not yield a significant difference. For the "uninstructed" groups, differences significant at or better than the 5 per cent level of confidence were obtained for *F* vs. *E* and *F* vs. *G* indicating positive transfer as having occurred not only in the opposable but also nonopposable initial-transfer situations.

E. DISCUSSION

1. *Differential Difficulty of Tasks*

The spatial relationships between stimulus and response in Tasks I and II were such that both the stimulus and response involved common axes, either horizontal or vertical, with direction given by either one of the colored lights. Because of this simplicity of spatial relationships, it was assumed that Tasks I and II would be much less difficult than Tasks III and IV where the stimulus occupied one axis, e.g., horizontal while the response the other axis, vertical. When it became clear that the results were reversing the expected group performances, the remaining subjects to be tested on either Task III or IV were asked at the end of the transfer task to give an account of how they performed the task in question. Many reports present a common operational method: The stimulus was simply seen as an axis in space and, depending upon the direction indicated, the correct response was either to the right or left or above or below the axis. Even during initial learning, many of these subjects did not look at the response panel when pushing the switches. Instead they "saw" the correct response or even projected themselves to be above, below, to the right or left of the stimulus axis. If, in Tasks III and IV as contrasted to I and II, the subject's perceptual operations are confined more to the stimulus panel and less to the response panel, less time per response would be required and more correct responses, hence, superior performance might be realized. There is, of course, the possibility that Tasks III and IV were the more challenging which may have been responsible for the superior performance.

2. *Effects of "Instructed" vs. "Uninstructed" Conditions*

The type of "instructions" given in this investigation is quite different from that of previous experiments which investigated the effects of particular principles applicable to the operation of the transfer task. The designed effect of pointing out beforehand all the correct responses for all the possible stimuli was to reduce initially the degree of directed attention required to learn or perform a relatively simple motor discriminative task. This task, moreover, was capable of being performed at a progressively lower level of directed attention as practice continued. In contrast to this designed effect was the requirement of greater initial directed attention which was to be met by the "uninstructed" subjects who had to "learn" the correct responses.

If proactive interference is, in part, an inverse function of the level of directed attention, then there should be demonstrated by the "uninstructed" groups a less degree of susceptibility to the greater interference evidenced by the "instructed" groups in the nonopposable initial-transfer situation. While such has been noted, the differences in degree of mastery attained by these two groups do not permit directed attention to be considered as the single determining cause.

3. *Determination of Transfer Sign*

The point of issue between the Wylie hypothesis of negative transfer and the hypothesis proposed by Gagné *et al.* concerns the condition which is the most productive of interference. Given physical identity of stimuli in the initial-transfer situation, the former hypothesis declares dissimilarity of response members to be responsible for that condition. Given identical stimuli and responses in the initial-transfer situation, the latter hypothesis submits as most productive of interference the condition of reversal of stimulus-response relationships. There is, however, more involved in this latter condition than the interference engendered by such reversal. A necessary inclusion is the factor of similarity-dissimilarity. In the example used by Gagné *et al.* (1) R_1 (right switch) is learned to S_1 (red light) and R_2 (left switch) to S_2 (green light) in the initial task. For the transfer task, the responses are reversed so that R_2 is learned to S_1 and R_1 to S_2. In relation to a given stimulus, e.g., S_1, response R_2 which must now be learned to that stimulus can be either similar or dissimilar to R_1 depending upon the degree of similarity between them. In the reversal tasks of this investigation, the similarity-dissimilarity dimension was represented by a perceivable aspect of similarity between the initial-transfer response members. The results of the investigation, in clearly demonstrating consistent superior transfer learning when initial and transfer response members are opposable (similar) and a strong possibility of negative transfer when these response members are nonopposable (dissimilar), support the Wylie hypothesis.

It is possible, of course, that a reversal of stimulus-response relationships may occasion a process similar to the hypothesized reciprocal inhibition of antagonistic reactions advanced by Osgood (7, 8). Yet, it has been seen that in the presence of an element of similarity between initial-transfer response members, such a possible inhibitory effect is of secondary importance. A like degree of importance could also be assigned to the effects of the greater amount of unlearning which might reasonably be claimed to occur.

The motor discriminative tasks of this investigation involved discrete responses capable of being organized spatially and made in relation to the discriminated spatial aspects of the stimuli. Accordingly, it is submitted that positive transfer, in a similar learning situation and where the stimulus-response relationships existing in a previous task are reversed in the subsequent task, will generally occur whenever there exists for the response members a perceived similarity such as given by communality of spatial axes.

There are many factors (5) which would occasion an exception to this statement. In particular there are two such important factors deserving of mention: (*a*) Number of mediated reversals (4). In this investigation, a correct response was accomplished by making a single movement in a single direction in relation to a single directional stimulus. For a correct response in the opposable transfer task, it was only necessary to make a single opposable movement. This task for this reason is one of simplicity and positive transfer is assured on the basis of similarity between response members. If the number of common axes or mediated reversals necessary to the accomplishment of a correct response is increased, the resulting greater complexity would act to reduce the amount of positive transfer. At some still greater degree of such complexity, it would be expected that negative rather than positive transfer would occur initially because of the interference *between* the different mediated reversals. (*b*) Habituation to interference. It has been generally agreed (5) and experimentally shown (12) that the greater the exposure to interference, the greater the degree of habituation to interference effects. If, therefore, the nature of the opposable initial-transfer situation is common to past behavior in which interference was experienced and habituation developed, then positive transfer is expected. Thus, when Hunter (3) reversed the maze which had previously been learned by human subjects, positive transfer was strikingly demonstrated. When, however, the particular transfer situation is uncommon to past experience, negative transfer is more likely to occur. Such was also reported by Hunter (3) for the rats which were required to make a simple discrimination reversal. This reported difference of positive and negative transfer occurring in the performances of humans and rats respectively cannot be explained solely on the basis of differences in ability. Using monkeys, Harlow (2) has demonstrated that not only does discrimination learning improve with the amount of such learning but that learning of discrimination reversal problems improves with the greater number of these problems learned.

Lastly, provided the required conditions of similarity and dissimilarity of stimulus and response members respectively, it is accordingly submitted that *negative* transfer is more likely to occur in other motor discriminative situations similar in nature to those of this investigation.

F. SUMMARY AND CONCLUSIONS

One hundred twenty-eight righthanded, male, basic Air Force trainees were assigned to eight different groups which permitted analysis of (*a*) differential difficulty of the four tasks made possible by the modification of *SAM* Self-Pacing Discrimination Reaction Time Test (CP611E), (*b*) effects of instruction upon proactive interference, and (*c*) differential motor transfer as a function of response similarity given by the communality of spatial axes in the case of rectilinearly opposable response members and response dissimilarity which was achieved by requiring the initial and transfer responses to be made upon different (orthogonal) axes. Before the initial and transfer tasks, there was indicated to four of the groups all of the correct responses for all of the possible stimuli. The remaining four groups were only told that they were to learn the correct response for each of the stimuli. All groups were given eight 2-minute trials on their respective initial tasks, followed by a 10-minute rest and then eight 2-minute trials on their respective transfer tasks. Scoring was in terms of correct and incorrect responses.

Conclusions derived are:

1. Superior performance resulted in those initial tasks where the spatial axes of stimulus and response were not in common but orthogonal.
2. The greater degree of directed attention and motivation which might be expected to occur when correct responses must be learned, in contrast to the case where they are pointed out beforehand, may be considered as factors responsible in reducing effect of proactive interference.
3. The differential motor transfer attributable to the similarity-dissimilarity of response members suggests that with identical stimulus members, negative transfer is more likely to occur when the motor response members are dissimilar.
4. Positive transfer, on the other hand, is more likely to result when the response members are similar and this also applies to the motor discrimination situation where the initial response and transfer response involve the same spatial axis which would result when stimulus-response relationships are reversed so that in the transfer task, the correct response is "opposite to" the correct response of the initial task.

REFERENCES

1. GAGNÉ, R. M., BAKER, K. E., & FOSTER, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67-79.
2. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51-65.
3. HUNTER, W. S. Habit interference in the white rat and in human subjects. *J. Comp. Psychol.*, 1922, **2**, 29-59.
4. JONES, F. E. Stimulus-response relationship as a dimension of task similarity-dissimilarity. Ph.D. Dissertation, Univ. Rochester, 1951.
5. MCGEOCH, J. A. The Psychology of Human Learning. New York, London, Toronto: Longmans, Green, 1946.
6. MELTON, A. W. (*Ed.*). Apparatus Tests. AAF Aviation Psychology Program Research Reports, No. 4. Washington: U. S. Government Printing Office, 1947.
7. OSGOOD, C. E. Meaningful similarity and interference in learning. *J. Exp. Psychol.*, 1946, **36**, 277-301.
8. ————. An investigation into the causes of retroactive interference. *J. Exp. Psychol.*, 1948, **38**, 132-154.
9. ROBINSON, E. S. The "similarity" factor in retroaction. *Amer. J. Psychol.*, 1927, **39**, 297-312.
10. SIIPOLA, E. M. The relation of transfer to similarity in habit-structure. *J. Exp. Psychol.*, 1941, **28**, 233-261.
11. SISSON, E. Retroactive inhibition: The influence of degree of associative value of original and interpolated lists. *J. Exp. Psychol.*, 1938, **22**, 573-580.
12. SPIETH, W., & LEWIS, D. The effects of alternating practice on the performance of two antagonistic motor tasks. Office of Naval Research, Tech. Report SDC 938-1-6.
13. WYLIE, H. H. An experimental study of transfer of response in the white rat. *Behav. Monog.*, 1919, **3**, No. 16.

USAF School of Aviation Medicine
Randolph Field, Texas

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 379-385.

AN APPLICATION OF THE ENGRAMMATIC CONCEPT*

Department of Psychology, Brooklyn College

IVAN D. LONDON[1]

The engrammatic concept, as developed in previous papers (1; 2; 3; 4, p. 147), implies both the formation of engram-complexes to accompany the learning process and their stabilization to accomplish retention. The engram-complex itself may be thought of as consisting of activated engram-units which are liable to inactivation and which are in spatially non-contiguous distribution throughout a neural network. These engram-units, while subject to spontaneous inactivation as previously noted, simultaneously undergo a restorative reactivation, thereby preventing the otherwise precipitous descent of the retention-curve to an asymptotic zero.

For purposes of visualization, the engram-complex may be thought of as bounded. But strictly speaking, there are no boundaries to an engram-complex other than those set by the neurones themselves. However, it is useful to think of the engrammatic boundary in approximation as closed surfaces dividing the much denser concentrations of activated engram-units from the considerably less dense. An engram-complex may then be visualized in approximation as a number of patterned bounded regions in functional interconnection within a neural network.

The molecular interpretation of the engram-unit leads to viewing the activated engram-unit as a complex molecule in one of its quasi-isomeric energy-states.[2] This activated molecule is always a candidate for spontaneous reversion to previous energy-levels and also for subsequent reactivation through restorative charging processes from outside it. This activated molecule may, in addition, be liable to further activational increase through assumption of a yet higher level of energy-state with corresponding accompaniment of the quasi-isomeric, "half-temporary" change of molecular structure that such increase implies.

---

*Received in the Editorial Office on June 12, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The writer gratefully acknowledges the support of the Inwood Institute, Upper Montclair, N. J. in behalf of research out of which this paper was developed.

[2] Isomeric molecules are composed of the same elements united in the same proportion by weight, but differing in some properties because of difference in structure. When a molecule is excited, its constituent atoms undergo temporary rearrangement. Hence, when a molecule is said to be capable of assuming quasi-isomeric energy-states, we mean that its distinct energy-levels correspond to significantly different and relatively stable configurations of the atoms constituting it (6, pp. 54-55).

Activation of an engram-complex can then involve any or several of the following in varying combination: (*a*) increase in engrammatic density, that is to say, increase in the number of activated engram-units within given bounded regions, (*b*) supplementary stepping up of activational or energy levels within given bounded regions as defined above, (*c*) expansion of bounded regions to include densities newly created through activation of neighboring potential engram-units, and (*d*) appearance of additional bounded regions to encompass newly formed clusters of activated engram-units.

The utilization of the engrammatic concept leads to derivations of ideal curves of retention for both overlearning (2) and non-overlearning (1). On assumption of the essentially molecular nature of the engram-unit, it leads also to the prediction that under certain conditions rate of forgetting may be shown to be in part a function of internal body temperature (3). In this paper it is proposed to apply the engrammatic concept to certain aspects of the conditioning process.

The formation of a conditioned response may be thought of as the establishment of a corresponding engram-complex. Let us assume that the initial process of conditioning a stimulus to a response organizes and marks out in approximation the boundaries of an engram-complex. Let us assume further that, until the conditioned response is firmly established, the engram-complex is in a state of partial consolidation only, that is to say, within the complex, whose approximate boundaries have been defined in the initial process of conditioning, there exist inactive engram-units potentially set for reinforcing recruitment into activated membership in the engram-complex. This recruitment is a process tending toward increased engram-unit activation and, hence, toward increasing consolidation of the designated engram-complex. In other words, within the bounds of the engram-complex there exist units, eligible for activation and awaiting recruitment to active participation as functioning constituents of the engram-complex.

Let us make two further assumptions for the conditioning process: (*a*) a stimulus of low magnitude marks out a smaller engram-complex, in terms of the number of engram-units defining consolidation, than a stimulus of higher magnitude; and (*b*) for a given stimulus the magnitude or strength of response is proportional to the "saturation" or "engrammatic density" of the consolidating engram-complex where saturation is relative to the number of engram-units taken as associated with consolidation. In view of available experimental evidence these are not unreasonable assumptions.

Let us limit ourselves to one activational level for the engram-unit. Let us, further, assign an effective rate of recruitment to counteract engram-unit inactivation which takes place concurrently.[3] Then, if as a descriptive approximation we take the rate of activation as proportional to the momentary number of available units potentially set for membership in the engram-complex, it would follow that the limiting magnitude or strength of a conditioned response is a function of the magnitude of the conditioned stimulus, provided, of course, that a physiological limit has not in the meantime intervened before the theoretical maximum has been obtained. In other words, in the process of conditioning a limit is reached for the magnitude or strength of a conditioned response regardless of the number of times a conditioned stimulus of given magnitude may be presented. To show this, we proceed as follows: Let the constant $P =$ number of activated engram-units of an ideally consolidated engram-complex corresponding to a stimulus of given magnitude or, stated otherwise, the number of units whose activation would not only set up, but consolidate the envisaged engram-complex. Let the variable s represent the number of presentations of the stimulus in the course of conditioning. Let the variable N be taken as the number of activated engram-units.

Then, the differential equation, describing the process of activation, is

$$\frac{dN}{ds} = k\,(P - N). \tag{1}$$

The solution of (1) is

$$N = P - ce^{-ks}, \tag{2}$$

where c is the constant of integration. Since ideally $N = 0$ when $s = 0$, $c = P$. Therefore,

$$N = P(1 - e^{-ks}). \tag{3}$$

If the constant $P' =$ number of activated engram-units in that consolidated engram-complex whose corresponding magnitude of stimulus is associated with a maximal response identical with that imposed by the physiological limit, then

$$n = p\,(1 - e^{-ks}), \tag{4}$$

where $n = N/P'$ and $p = .P/P'$. In other words, the physiological limit is taken as the theoretical base for percentage calculations.

When the conditioned stimulus is weak or of moderate strength, p is less than 100 per cent. From (4) it then follows that $n = f(s)$ is asymptotic to a line parallel to, but below, the physiological limit and that the associated maximal conditioned response is likewise below the absolute maximal allowed by that limit. If the conditioned stimulus is of considerable magnitude, we surmise that p is greater than 100 per cent. Again from (4) it follows that $n = f(s)$ is asymptotic to a line parallel to, but above, the physiological limit and that the associated response exhibits its maximum much sooner than if p is less than 100 per cent. In the case of even intenser stimuli, $n = f(s)$ should appear linearized. These deductions are ideally illustrated with an assigned value of k equal to ½ in the accompanying figure where magnitude of response is taken as proportional to percentage of activated engram-units and the number of engram-units corresponding to the physiological limit is taken as base (Figure 1).

---

[3] Since engram-units are subject to a spontaneous inactivation in the course of recruitment, a suitable rate of effective recruitment must be postulated to counteract that of inactivation.

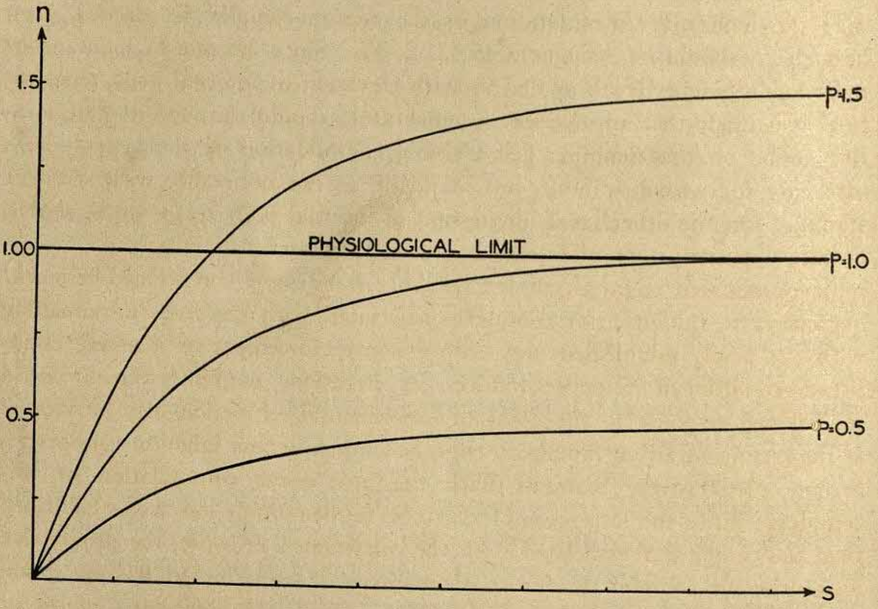


FIGURE 1

As long as the physiological limit is not exceeded, the maximal conditioned response can be directly used to yield an estimate of the relative "size" of the corresponding consolidated engram-complex. Where the physiological limit has been exceeded, the existence of an asymptotic $n = f(s)$ can be demonstrated indirectly by the readiness with which experimental extinction can be induced. Thus, it is in accordance with laboratory experience that those engram-complexes for which p is greater than 100 per cent should be proportionately more resistant to extinction than those for which p is less than 100 per cent.

If we introduce the factor of activation beyond a first level, the resulting picture becomes more complicated. However, any major activation to higher levels would seem most likely to occur with stimuli of larger magnitudes and would thus manifest itself in a more stubborn resistance to extinction. In the latter contingency, the relative "size" of the engram-complex cannot be accurately gauged, since a considerable proportion of its engram-units are superactivated, that is to say, activated beyond the first energy level with the result that spontaneous reversions from higher levels must run step-wise through a sequence of successively lower levels before reaching the level corresponding to inactivation.

If the molecular interpretation of the engram-complex is adopted, then heat effects should be demonstrable (1, 2, 3). Since increase of spontaneous reversion of energy-levels is tied in with elevation of internal body temperature, it is likely that an increase in temperature should, through its deteriorative action on the complex, delay visible consolidation of the latter—more markedly for weaker stimuli, less markedly or not noticeably with stronger stimuli. On the other hand, depression of internal body temperature should operate oppositely: consolidation should be facilitated.[4]

A consolidated engram-complex not only stabilizes a conditioned response, but comes to inhibit prior complexes associated with responses incompatible with the newly established one. Since nonreinforcement of a newly established conditioned response induces its negation, nonreinforcement of a conditioned response to the point of experimental extinction may be viewed as the setting up of an inhibitory engram-complex whose inhibitory properties become progressively manifest with the progressive consolidation of this complex. Since the latter complex (to be termed *B*) is less well-established than the complex *A* associated with the conditioned response, the progressive dissolution of complex *B* due to spontaneous inactivation of its units soon frees the previously established complex *A* from the inhibitory effects of complex *B* which the extinctual process had set up. In other words, we have spontaneous recovery.

Since the inhibitory complex *B* inactivates progressively with time in accordance with a previously derived law (1), it manages to maintain for some time a fractional existence even when complex *A* has been freed of its inhibitory effects. This fractional base makes it easier for a second run of experimental extinction to assert itself. A more consolidated complex *B* results. This means a reduction of spontaneous recovery, since degree of recovery is inversely related to degree of integrity of complex *B*. The successive repetition of the above argument leads to the correlation of continual decrease in degree of spontaneous recovery with increasing consolidation of the inhibitory complex *B*. However, if complex *A* has been well-established, a consolidated complex *B* that has not been as overlearned or superactivated may be counted upon to inactivate progressively until a point is reached where its inhibitory effects are no longer such as to be able to mask the presence of complex *A*. Spontaneous recovery will again take place. All this is in accord with the evidence of the laboratory.

---

[4] Considerations of engrammatic activation as a thermal function have not been overlooked. The meager evidence available indicates that rate of learning may be thermally dependent (**5**). However, in this discussion the assumption is made that activational rate is much less thermally affected than inactivational rate. Where this is not true, either masking or a variety of complicating effects may be anticipated.

If the molecular interpretation of the engram-complex is assumed, the effects of internal body temperature may be demonstrable. A moment's reflection shows that the younger, less established complex *B* should deteriorate more readily with elevated internal temperature than should complex *A*. Thus, we may conclude that spontaneous recovery will be more readily and vigorously manifested with increase of internal body temperature. This, at first, seems to contradict the deduction advanced in a previous paper that rate of forgetting increases with elevation of internal body temperature (1). However, it is seen that spontaneous recovery is a function of the "forgetting" of extinction, that is to say, of the degree of dissolution of the inhibitory complex. Thus, on elevation of internal temperature the inhibiting complex *B* suffers greater loss of integrity than the inhibited complex *A* with consequent quickened manifestation of the latter. On the other hand, since decrease of internal temperature retards dissolution of the engram-complex, for this contingency a contrary effect may be anticipated—relative delay and enfeeblement of spontaneous recovery.

Several deductions have been presented in this paper on the basis of an engrammatic theory. The question as to whether every effect deduced from this version of engrammatic theory is experimentally demonstrable remains a pertinent one. For example, if the learning process involves microstructural changes, as the writer believes is the case, if biophysical and biochemical elements are involved, then it must seem that the factor of internal heat is one of some importance. But whether the effect of internal heat on learning and retention is directly demonstrable, still is a matter for conclusive determination.

REFERENCES

1. LONDON, I. D. An ideal equation derived for a class of forgetting curves. *Psychol. Rev.*, 1950, **57**, 295-302.
2. ———. An ideal equation derived for overlearning. *Psychol. Rev.*, 1951, **58**, 54-59.
3. ———. A theoretical relation between rate of forgetting and internal body temperature. *J. Gen. Psychol.*, 1952, **46**, 151-157.
4. ———. Quantum biology and psychology. *J. Gen. Psychol.*, 1952, **46**, 123-149.
5. PIÉRON, H. Sur l'accélération, avec la température, de l'évolution des processus mnémoniques. *C. R. Séances Soc. Biol.*, 1941, **135**, 631-633.
6. SCHRÖDINGER, E. What Is Life? New York: Macmillan, 1945.

Department of Psychology
Brooklyn College
Bedford Avenue & Avenue H
Brooklyn 10, New York

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 387-392.

A FACTOR ANALYSIS OF THE 48 SZONDI PICTURES*

U. S. Naval Personnel Research Unit, San Diego, California

LEONARD V. GORDON[1]

A. INTRODUCTION

The Szondi test consists of 48 photographs of the faces of mental patients. These pictures are divided into six sets, every set containing one member from each of eight psychiatric categories. The subject taking the Szondi selects, from each set, two pictures that he likes and two that he dislikes. The subject's scores on the subscales are derived directly from the psychiatric categories to which the selected pictures belong.

The use of psychiatric diagnoses for scoring categories is based on Szondi's (4) theory that an individual's response to a picture will be determined by his propensity toward the particular disorder of the patient pictured, lying latent in his own genes. These scoring categories have been found to lack reliability by both Gordon and Lambert (2) and Lubin and Malloy (3) who have shown that, for all categories, responses to one picture within a category is no more related to responses to other pictures within the same category, than to pictures in the test as a whole.

Gordon and Lambert, however, did find that the number of significant relationships among responses to Szondi pictures was well above chance expectancy. The present factor analysis was undertaken to account for these relationships in terms of independent dimensions underlying the liking or disliking of Szondi pictures.

B. PROCEDURE

The Szondi pictures were made into a film strip, with the sets appearing in the prescribed order, and the pictures randomized within each set. The pictures were presented to members of seven summer school classes at an Eastern university. After a trial run to accustom the subjects to the type of picture comprising the test, they were asked to indicate "like" or "dislike" for each picture. This modified method of presentation was used, rather than the standard procedure, in order to avoid ipsative scores, and to obtain the most valid reflection of interrelationships among all the pictures.

---

*Received in the Editorial Office on June 13, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The opinions expressed are those of the author and are not to be construed as being official or in any way representative of the U. S. Navy.

The tetrachoric correlations between pictures were obtained from the responses of 118 male and 82 female subjects, with the aid of the Chesire, Saffir, and Thurstone computing diagrams (1). The correlation matrix[2] was factored by a modified (6, 7) Thurstone Multiple Group Factor Analysis (5), with 13 factors extracted. Since the purpose of the study was to determine independent dimensions underlying the liking or disliking of the pictures, orthogonality of the axes was maintained. Toward the end of the rotational process, the factors were rotated for meaningfulness. The final orthogonal loadings of the 48 pictures on the 13 factors are presented in Table 2.

C. IDENTIFICATION OF THE FACTORS

Following are descriptions of the 13 factors based on pictures having significant (.31 or over) loadings on each. These pictures are listed, by factor, in Table 3. Since, in a number of the Szondi pictures, characteristics such as age, sex, mood are unclear, independent judgments were obtained from 31 subjects of such characteristics for each person pictured.

Factor *A* contains eight pictures with significant loadings, all of which are of clean cut, normal appearing males. They represent the more conventional, modern looking members of the Szondi group, and include those who are judged to be the younger men. VIh is the only person who is ambiguous as to sex. Five of the six homosexuals in the Szondi test are represented in this factor.

Factor *B* was hypothesized to represent femaleness during the later part of the rotational process, since all pictures judged to be female had significant or almost significant loadings on it. The particular quality of femininity represented in this factor is somewhat difficult to appreciate, since the individuals with the three highest loadings were judged to be male by a fair minority of the judges. The disagreements regarding the sex of many patients, whom the majority judged to be female, were probably occasioned to a large extent by their hair styling, which is masculine by present day standards, as well as the patient's wearing what appears to be institutional uniforms. The very high agreement as to the sex of the patients whom the majority judged to be male, was undoubtedly due to the bearded or distinctly unshaven appearance of many of these patients, as well as their

---

[2] To save printing costs, Table 1 containing the correlation matrix has been deposited with the American Documentary Institute. Order document No. 4040 from American Documentary Institute, 1719 N. St. N. W., Washington 6, D. C., remitting $1.25 for 35 mm. microfilm or $1.25 for 6 by 8 in. photocopies.

TABLE 2
ORTHOGONAL ROTATED FACTOR LOADINGS* AND COMMUNALITIES

Picture	A	B	C	D	E	F	G	H	I	J	K	L	M	h^2
Ih	13	—05	06	11	—20	22	37	00	10	—01	—13	—07	—07	30
Ie	06	07	48	—06	02	—13	—03	12	—09	25	02	00	—11	36
Ip	—17	02	39	24	02	19	12	21	—06	09	49	13	—11	61
Ik	—03	19	01	27	70	20	08	11	—09	02	—10	00	20	72
Id	20	53	—15	06	00	01	—17	32	04	—10	05	01	15	52
Ihy	02	14	27	47	—14	—03	—09	10	14	24	15	11	09	47
Is	12	—04	06	12	—03	72	17	10	12	—22	—01	00	04	66
Im	—18	33	12	05	07	37	14	—07	39	00	—12	—10	—02	50
IId	01	38	19	—17	—03	—18	09	28	75	—22	—01	04	—08	95
IIh	54	24	—20	12	—06	06	—01	13	12	—08	—07	17	—15	50
IIhy	15	37	13	00	66	06	04	06	11	—04	—04	24	13	71
IIs	44	04	09	—01	04	05	03	10	25	—06	—33	13	73	94
IIe	12	50	—01	19	15	—01	06	32	03	06	—10	17	—14	49
IIp	—07	—15	14	65	33	—13	17	12	07	00	—01	—01	07	65
IIk	—01	51	12	—05	—18	—01	—02	58	01	—05	22	—06	07	71
IIm	—10	13	07	19	—23	06	—14	—05	17	10	03	30	—11	29
IIId	—13	44	04	20	—18	05	12	34	—05	07	12	—12	—14	47
IIIm	43	00	01	—09	04	32	13	14	01	—16	00	79	00	98
IIIk	—10	31	15	16	72	—04	06	12	—04	15	11	16	—21	80
IIIhy	06	—02	—07	08	32	—08	03	01	—14	—12	87	—17	19	98
IIIs	22	14	11	05	—13	15	—09	—05	08	—04	16	—13	10	19
IIIe	—02	15	64	16	—14	12	—16	17	19	34	16	21	—02	77
IIIh	67	12	01	—07	—01	—15	50	—02	02	20	—06	20	—12	84
IIIp	—08	34	06	41	—24	04	—04	09	17	15	12	12	—14	46
IVhy	01	58	—15	—14	—13	13	08	16	—08	13	—04	15	15	51
IVk	—13	11	25	23	19	04	52	02	02	18	07	29	19	61
IVh	82	12	—08	—04	15	14	00	—05	11	03	—11	11	—08	78
IVd	08	06	00	26	10	—05	09	14	—19	13	—01	65	02	59
IVe	—02	—03	—04	63	23	—16	07	08	04	09	—02	24	—11	57
IVp	13	09	04	53	—11	07	58	21	00	05	—16	—18	03	77
IVm	07	31	—11	24	07	06	01	—02	—05	—09	—11	—13	00	22
IVs	22	04	—08	34	36	07	—13	—07	17	—09	01	05	12	38
Vs	04	20	06	18	01	21	14	02	13	14	07	—20	14	24
Vh	36	01	10	—11	—15	13	06	16	16	—16	—23	—23	04	38
Vhy	22	38	—19	03	11	31	14	00	08	06	12	02	01	37
Ve	53	05	07	—33	12	—01	09	11	—07	01	08	—05	57	77
Vp	06	12	24	26	—08	02	09	15	15	10	07	15	13	26
Vk	04	50	13	08	—07	03	21	—24	—06	54	11	—09	—02	70
Vd	18	36	—05	03	01	04	15	—17	32	12	25	15	01	42
Vm	—16	08	09	06	14	19	14	—07	33	—08	00	—04	01	24
VId	18	44	—18	20	—09	02	00	—14	72	—02	—12	04	04	86
VIe	02	38	—01	03	22	36	—01	09	09	00	—18	—05	04	38
VIk	11	03	43	03	18	53	10	07	—11	17	—13	20	11	64
VIs	12	—09	28	30	10	—05	—04	17	—04	33	—11	—01	16	38
VIm	01	22	24	50	07	03	—11	04	02	21	61	—05	08	80
VIhy	04	32	—05	33	—03	—08	10	10	—13	48	03	—02	—02	49
VIp	06	19	48	20	28	—10	03	—05	14	52	—03	01	11	70
VIh	53	38	—30	09	01	19	11	06	10	—08	—09	—09	28	69

*Decimal points omitted.

TABLE 3
SZONDI PICTURES HAVING SIGNIFICANT LOADINGS IN EACH FACTOR

Factor	Pictures and loadings*
A	IVh(82) IIIh(67) IIh(54) Ve(53) VIh(53) IIs(44) IIIm(43) Vh(36)
B	IVhy(58) Id(53) IIk(51) IIe(50) Vk(50) IIId(44) VId(44) VIe(38) VIh(38) IId(38) Vhy(38) IIhy(37) Vd(36) IIIp(34) Im(33) VIhy (32) IIIk(31) IVm(31)
C	IIIe(64) VIp(48) Ie(48) VIk(43) Ip(39)
D	IIp(65) IVe(63) IVp(53) VIm(50) Ihy(47) IIIp(41) IVs(34) VIhy(33) Ve(—33)
E	IIIk(72) Ik(70) IIhy(66) IVs(36) IIp(33) IIIhy (32)
F	Is(72) VIk(53) Im(37) VIe(36) IIIm(32) Vhy(31)
G	IVp(58) IVk(52) IIIh(50) Ih(37)
H	IIk(58) IIId(34) Id(32) IIe(32)
I	IId(75) VId(72) Im(39) Vm(33) Vd(32)
J	Vk(54) VIp(52) VIhy(48) IIIe(34) VIs(33)
K	IIIhy(87) VIm(61) Ip(49) IIs(—33)
L	IIIm(79) IVd(65)
M	IIs(73) Ve(57)

*Decimal points omitted.

wearing either modern-type attire or having the appearance of being unattired from the waist up.

Factor *C*, like Factor *A*, also contains younger males. However, there is a grimy appearance about the members of the present group, in contrast to the neat, clean cut appearance of the members of Factor *A*. The lighting in the pictures of Factor *A* is relatively homogeneous, while there is a predominance of shadow, particularly about the eyes, of the individuals in the present group. This shadowing may account, at least in part, for their grimy appearance. While the individuals in Factor *A* appear to be conventional young men, those in Factor *C* are perhaps best designated as "bums."

Factor *D* contains, as individuals with the highest loadings, all of the serious, older, bearded males. IIIp is the oldest female by a full generation, while Ve, with a negative loading, is a smiling or leering younger person. This factor appears to represent some combination of seriousness and old age.

Factor *E* has, with the highest loadings, those persons who were judged to be the most peculiar members of the Szondi family, and together with IIIhy, the most disliked. These three individuals and IIp are the only persons with distorted facial expressions.

Factor *F* contains individuals whose most common characteristic is a smiling countenance. However, since several other persons not represented on this factor are judged to be smiling, identification of this factor is not entirely clear.

Factor *G* is made up of individuals who are judged to be somewhat older, on the average, than those in Factor *A*, but are likewise clean cut and relatively normal looking. This group presents more of a "professorial" appearance than any of the other groups.

Factor *H* has four persons with significant loadings. They are all depressed-looking females, two of whom are diagnosed as depressives. It may be noted that three of the four are wearing identical apparel. The only other two members of the Szondi group wearing this type of dress pattern are two of the peculiar looking individuals in Factor *E*. The type of dress worn may contribute to some extent to the common factor variance.

Factor *I* has, as individuals with significant loadings, manics and depressives. In the average judgments of the mood of these people, there was general agreement that the two manics were happy people, VId was placed on the unhappy side of the "neither" position, while the other two depressives were placed on the happy side of the "neither" position. Thus, it cannot be said that the depressives in this group mirror their diagnosis in their facial expressions.

Factor *J* is difficult to identify. The individuals with significant loadings on this factor have in common a stare that cannot be said to be absent from the other Szondi pictures, but which is still the most obvious common characteristic.

Factor *K* contains three persons with significant, positive loadings. They are judged to be rather sombre, morbid-looking males, while IIs, with a significant negative loading is decidedly cheerful in appearance.

Factor *L* is obviously a doublet, where at first glance the two pictures appear to be of the same person, with different expressions in each picture.

Factor *M* also appears to be a doublet composed of two similar-looking, corpulent, young men.

D. Discussion

The results of the factor analysis indicate that an individual's liking or disliking of a Szondi picture is to a large extent related to such identifiable characteristics of the patient pictured as age, sex, mood, conventionality of appearance, etc. Thus, it is possible that attitudes towards particular characteristics of people, which are observable in or inferable from photographs, may be measured by the Szondi pictures.

In its standard administration, the test probably cannot be scored to yield reliable measures of these attitudes, since there is little similarity of factor content between sets and a wide range of popularity or preference value occurs within sets. However, some modified method of picture presentation, such as that used in the present study or "forced-choice" presentation might be expected to give somewhat reliable scores.

Considerable research has been performed on the measurement of attitudes related to those disclosed in the present study; e.g., toward the opposite sex, parental authority, conventional standards. A relatively indirect and somewhat disguised approach, such as the technique of picture presentation, may be of value in such measurement.

REFERENCES

1. CHESIRE, L., SAFFIR, M., & THURSTONE, L. L. Computing Diagrams for the Tetrachoric Correlation Coefficient. Chicago: Univ. Chicago Bookstore, 1933.
2. GORDON, L. V., & LAMBERT, E. J. The internal consistency of the Szondi "factors." *J. Soc. Psychol.,* (in press).
3. LUBIN, A., & MALLOY, M. An empirical test of some assumptions underlying the Szondi test. *J. Abn. & Soc. Psychol.,* 1951, **46**, 480-484.
4. SZONDI, L. Experimentelle Triebdiagnostick. Bern: Huber, 1947.
5. THURSTONE, L. L. Multiple Factor Analysis. Chicago: Univ. Chicago Press, 1947.
6. WHERRY, R. J., BROGDEN, H. E., & GAYLORD, R. H. Wherry-Brogden-Gaylord Method of Factor Analysis. Personnel Research Section, Adjutant General's Office, Department of the Army, Washington, D. C. (unpublished).
7. WHERRY, R. J. A new iterative method for correcting erroneous communality estimates in factor analysis. *Psychometrika,* 1949, **14**, 231-241.

U. S. Naval Personnel Research Unit
Classification Research Branch
San Diego 52, California

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 393-400.

THE AIMS OF PREVENTIVE PSYCHIATRY AND "CAUSALITY" AS A PERSONALITY PATTERN*

Preventive Psychiatry Project, Child Welfare Research Station, State University of Iowa

EUGENE E. LEVITT AND RALPH H. OJEMANN[1]

A. THE AIMS OF PREVENTIVE PSYCHIATRY

Psychiatry may be generally defined as the specialized discipline which treats mental disorders. Preventive psychiatry broadly speaking, is aimed at the prevention of mental disorders and the promotion of emotional maturity through a systematic attack upon the causes of such disorders. More complete designations of the purposes of an organized preventive psychiatry program have been presented by Aberle (1) and by the Group for the Advancement of Psychiatry (17). Four aspects of these formulations apply directly to the Preventive Psychiatry Project at the State University of Iowa:

1. To carry out research on problems of the relationship of disturbances in emotional well-being to the social system in which they occur, particularly family and school areas.
2. To operate as a team with a broad social science approach.
3. To set up collaborative relationships with such policy-making groups as schools and churches so as to understand their problems and to be able to offer help with general questions of preventive psychiatry.
4. To design learning experiences for the growing personality which will allay or prevent "disturbances in emotional well-being." These experiences are designed, in accordance with their purposes, to effect modifications in human behavior and attitudes rather than to inculcate mere information as the ordinary educational use of the word "learning" often seems to imply.

A program of this type, which includes a large variety of individual researches, requires a unifying concept both to provide an origin for research and to effectively focus the experimental work.

The unifying concept of our project is *"causality."* "Causal" is roughly

---

*Received in the Editorial Office on June 13, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] The preparation of this paper was supported by Research Grant MH-301 from the National Institute of Mental Health of the National Institutes of Health, U. S. Public Health Service. The authors gratefully acknowledge the assistance of Drs. Boyd R. McCandless, Alfred Castaneda, and Seymour L. Zelen; Miss Marjorie A. Moore, Mrs. Margaret Ehmke, and Mr. Derek Walton.

synonymous with "dynamic," "motivational," or "analytical" as these terms are commonly used in contemporary psychology. To behave causally is to deal with inter-personal phenomena in a way which takes account of the factors underlying or causing behavior (11). The antonym for causal is "surface," that is, an egocentric or stereotyped approach to human interaction which ignores behavior dynamics. This crude formulation will be elaborated in subsequent pages.

The selection of causality as the integrating idea of the project was based on a preliminary hypothesis, which relates causality to mental health. We paraphrase this hypothesis from Ojemann (12): Causality is associated in humans of a given age with psychological adjustment, emotional well-being, psychological health, and similar concepts as these terms are generally understood in psychology and psychiatry.

The investigation of the concept of causality as a behavioral variable is thus in accord with the first statement of the general purposes of preventive psychiatry (see p. 393).

A number of detailed anecdotal accounts of the nature of causality as a pattern of behavior have been published during the past 15 years by the present project director and his students (7, 13, 16). Up to this time, no formal definition of this concept has appeared in the literature. The primary purpose of this paper is to furnish this definition, and to show how it is used to guide research within the project.

B. THE DEFINITION OF CAUSALITY

1. *Some Assumptions*

Before presenting the definition of causality, there are certain assumptions correlative to the definition which should be made explicit.

1. Causality as we see it is a complex pattern of behavior. Any such pattern involves the individual's entire frame of reference, his perceptual tendencies, his motivational and emotional facets, and cognitive areas like beliefs and opinions. Theoretically then, causality can be conceived of as a complex personality syndrome, with certain behavioral consequences which we shall attempt to predict.

2. Causal behavior is in part a function of development of cognitive processes, especially the learning of the complex symbolic behavior necessary for adaptation in our society. Different types of behavior which are classifiable as causal can be expected to appear at various age levels as a function of intellectual development. The definition of causality which

we will present is applicable to adult behavior, and to the behavior of adolescents. Its appropriateness for the very early age levels is uncertain since many of the behaviors subsumed within the definition are beyond the available repertoire of the small child. The definition may be adjusted to age levels after empirical investigation.

3. In general, the training and experiences provided for the growing child by parents, other influential adults, and social institutions in our culture are likely to be non-causal or surface. Evidence presented by Ojemann (9, 12) supports this contention. Further examination of existing degrees of causality at various educational levels is being planned by the project.

2. *Methodological Observations*[2]

1. As we have pointed out, causality may be conceptualized both as a pattern of behavior, and as a personality characteristic. Basically, the former construction is our primary one, since we are most concerned with effecting modifications of behavior. However, as we will hypothesize in the body of the definition, causal behavior depends, for its specific form, on environmental circumstances. It cannot be stipulated *in vacuo* except in the vaguest and broadest terms, terms which would have little if any empirical meaning. Therefore, the definition will be stated in terms of certain underlying cognitive and personality characteristics, from which behavior logically appropriate to various situations can be specified, and subjected to experimental analysis.

2. In order to have wide experimental application as in program research, a definition must be relatively abstract, grouping a range of "definitional chains" or abstractions.

3. The purpose of such a broad abstraction is to provide a general framework into which a number of experimental designs can be fitted. This is a necessary requisite for program research since the adoption of any single "operational definition" would seriously limit the scope of research.

4. Experimental designs stemming from the original abstraction constitute operational definitions for different phases of the abstract definition.

5. Support of the null hypothesis in an individual experiment should result in consequent amendation of the abstract definition. When specific operations fail to provide predictive results, some facet of the original abstraction is deemed inappropriate.

---

[2] The project is not limited to one, specific *formal* theoretical or methodological approach. Staff members are free to use any theory or method which can be demonstrated to be useful in the development of the project. The basic methodology set forth in this paper has thus far been found to satisfy this criterion.

It should be clear at this point that the abstract, central definition is actually a complex hypothesis. This hypothesis is stated in terms which are themselves not immediately capable of experimental verification. The operational definitions deduced from the abstraction can be analyzed experimentally. In this way, the hypothesis is subject to objective test.

The predictive operational definition attests to the usefulness of the construct from which it was derived. If there is no prediction, the construct is no longer of value, and is discarded. This procedure may be unnecessary for an individual experiment, but it has merit for program research since it permits a wide investigation of a particular concept.

3. *The Definition*

Causality is behavior of some kind. The particular form of the behavior will depend upon the relationship between the condition of the individual at the time of observation, and the circumstances in which he is placed. Causal behavior may take the form of withdrawal, aggression, compliance, resistance, and so on, depending upon these conditions. The possible structurings of the situation are infinite; it is not necessary to specify any structurings in this definition. Specification is, however, required in the investigation of a specific hypothesis about some phase of causality. As experimental evidence accumulates increasing specification can occur.

The condition of the individual at any point in time depends in part upon the nature of the situation (which includes the individual as a person). However, it is possible to hypothesize certain relatively constant conditions of the individual which underly the "momentary states," and are of special significance in determining their character. These conditions may be termed "behavior dispositions" for want of more precise terminology. The primary behavior dispositions are:

1. Knowledge of the existence of causes for human behavior, both of one's own behavior and that of others.

2. Knowledge that the underlying factors of behavior are often complex and interacting.

3. The ability to assume alternate multiple explanations of behavior; the awareness that different motivations may be behind the same behavior in different persons, and that the same underlying causes may produce different behaviors in different persons.

4. A disposition leading to consistent attempts to "see things from the viewpoint of others," that is, to identify with those with whom one inter-

acts. This is the most primitive form of investigation into the causes of behavior in another individual.

5. The realization that one's behavior will have effects upon the behavior of others.

6. Suspending of judgment of another's behavior until sufficient, logical information for such judgment or evaluation is available, i.e., until one particular hypothesis concerning the causes of that behavior (see 3 above) or the effects of behavior has achieved a maximum probability of being a true explanation. A "scientific" readiness to modify even this hypothesis in the face of inconsistent events is implied here.

We hypothesize further that these behavior dispositions fall into a logical hierarchy, in which a particular disposition cannot manifest itself in an individual unless the preceding dispositions are present. For example, a person cannot know that the causal factors of behavior are complex unless he is first aware that such causes exist. He cannot assume alternate multiple explanations unless he knows that behavior has causes, and that these causes are often complex, and so on. It is possible that the appearance of succeeding dispositions is a function of intellectual development as noted earlier in this paper.

Potentially then, an ordered scale for the measurement of causality exists, with each scale step defined. An individual manifesting evidence of three behavior dispositions is *ipso facto* more causal than one with only two dispositions. An individual manifesting all six dispositions has achieved maximum causalness.

Like all other points made in the definition, the hierarchical nature of causality must be subject to experimental verification.

The primary behavior dispositions of causality appear to be phrased in terms of cognitive capacities. However, these capacities do not usually exist in isolation from certain more obviously motivational conditions. Below are listed a number of such conditions which we hypothesize are linked with the primary dispositions. They are annotated in the construct terminology in which they usually appear in the literature of psychology.

1. *Flexibility,* i.e., non-rigidity of personality organization, following the definition of rigidity offered by Rokeach (15): "Rigidity . . . (is) defined as the inability to change one's set when the objective conditions demand it."

2. *Tolerance of ambiguity.* An individual who is tolerant of ambiguity lacks the "tendency to resort to black-white solutions, to arrive at premature closure as to valuative aspects, often at the neglect of reality, and to

seek for unqualified and unambiguous over-all acceptance and rejection of other people" (5).

3. *Democratic in attitude and belief,* that is, characterized by "affectionate, basically equalitarian, and permissive interpersonal relationships" (2).

4. *Empathy.* This is taken to mean "the ability to transpose oneself into the thinking, feeling, and acting of another" (4).

5. *Self-acceptance.* A self-accepting person is one who possesses insight into his own motivations and dynamics, who is aware of his limitations and drawbacks as well as of his capacities, and who is able to accept himself for what he is, and to be neither gravely distressed nor unduly motivated by his shortcomings. The importance of the self-concept for psychotherapeutic technique has received recent attention among clinicians. For a recent attempt to measure the self-concept see the work of Bugental and Zelen (3).

6. *Emotional maturity.* This concept is a devious one. A review of its various applications in psychology and psychiatry reveals a wide divergence of meaning. In view of this confusion, it is necessary to specify here that emotional maturity will be operationally defined by certain scales, and will not be considered to have surplus meaning beyond the defining measures.

An analysis of the primary behavior dispositions of causality indicates that what we might call "causal thinking" is roughly analogous to what the therapist labels "insight." Bringing an individual to the point of clear, insightful thinking concerning his own personality dynamics and his interrelationships with others is one of the classical mechanisms of psychotherapy. The argument for insight as a therapeutic device is convincing: A person cannot be expected to give up his present (neurotic) pattern of existence unless it becomes clear to him that the maintenance of this pattern is not in his own best interests. He cannot arrive at this realization unless he reorients his perception of himself and others. Such a reorientation is called insight.

It has been pointed out that insight may be a painful thing for an emotionally disturbed person, and that he may hence resist it strongly. One of the aims of preventive psychiatry is to foster this insight *before* behavior and attitude patterns have developed into serious maladjustment, that is, to engender insight into behaviors including those which might become neurotic if left without modification. This is the reason that the concept of causality as a personality and behavioral dimension concerns preventive psychiatry. Our definition clearly shows that we consider insightfulness to be a logical concomitant of emotional health. Contrariwise a rigid, stereotyped approach to human relations often breeds many personal problems.

When we have verified the hypotheses concerning the structure of causality, we can proceed to devise planned experiences designed to develop causality in growing personalities. An account of the nature, operation, and results of such programs has been presented elsewhere (10). There is evidence that skillfully designed and applied programs can effect significant reductions in personally-experienced conflict (8), changes in social attitude (6), and increases in pupil growth when applied to teachers (14). There is no reason to suppose that with increased efficacy of design, further changes cannot be effected. We hope eventually to apply our findings to the re-education of parents, the development of school curricula, the training of teachers, and the preparation of standard course textbooks.

REFERENCES

1. ABERLE, D. F. Introducing preventive psychiatry into a community. *Hum. Org.*, 1950, **9**, 5-9.
2. ADORNO, T. W., FRENKEL-BRUNSWIK, E., LEVINSON, D. J., & SANFORD, R. N. The Authoritarian Personality. New York: Harper, 1950. Pp. 990.
3. BUGENTAL, J. F. T., & ZELEN, S. L. Investigations into the "Self-Concept": I. The W. A. Y. technique. *J. Personal.*, 1950, **18**, 483-498.
4. DYMOND, R. F. Personality and empathy. *J. Consult. Psychol.*, 1950, **14**, 343-350.
5. FRENKEL-BRUNSWIK, E. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Personal.*, 1949, **18**, 108-143.
6. LINKS, E. A comparison of two methods of using radio for changing certain social attitudes. Unpublished Master's thesis, State University of Iowa, 1945.
7. MCNEIL, B. Development at the youth level of a conception of the causes of behavior and the effectiveness of a learning program in this area. *J. Exp. Educ.*, 1944, **13**, 81-85.
8. MORGAN, M. I., & OJEMANN, R. H. The effect of a learning program designed to assist youth in an understanding of behavior and its development. *Child Dev.*, 1942, **13**, 181-194.
9. OJEMANN, R. H. The effect on the child's development of changes in cultural influences. *J. Educ. Res.*, 1946, **40**, 258-270.
10. ———. Research in planned learning programs and the science of behavior. *J. Educ. Res.*, 1948, **42**, 96-104.
11. ———. An integrated plan for education in human relations and mental health. *J. Sch. Hlth.*, 1950, **20**, 99-106.
12. ———. The purpose, problems, and promise of preventive psychiatry. Paper presented before the Society for Research in Child Development, Yellow Springs, Ohio, 1953.
13. OJEMANN, R. H., & BYRNES, D. The teacher's reaction to child behavior. Publication No. 1, Integrated Plan for Teacher Human Relations and Mental Health, Preventive Psychiatry Project, State University of Iowa, 1949.
14. OJEMANN, R. H., & WILKINSON, F. R. The effect on pupil growth of an increase in teacher's understanding of pupil behavior. *J. Exp. Educ.*, 1939, **9**, 143-147.
15. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. Abn. & Soc. Psychol.*, 1948, **43**, 259-278.

16. STILES, F. S. Developing an understanding of human behavior at the elementary school level. *J. Educ. Res.*, 1950, **43**, 516-524.

17. Promotion of mental health in the primary and secondary schools: An evaluation of four projects. Report No. 18, Committee on Preventive Psychiatry of the Group for the Advancement of Psychiatry, January, 1951.

Preventive Psychiatry Project
Child Welfare Research Station
State University of Iowa
Iowa City, Iowa

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 401-408.

THE PERCEPTION OF SIZE OR DISTANCE BASED ON TACTILE AND KINESTHETIC DATA*

Department of Psychology, Michigan State College

S. HOWARD BARTLEY

A. PROBLEM

A blindfolded individual who is presented an object for tactile exploration, usually, if not always, attempts to visualize the object from the scant information obtained by exploring it. This is taken by most theorists simply to mean that the observer is merely trying to aid the tactual sense by *adding* visual imagery. The implication, if not their actual assertions, has it that an observer can tell the size of an object without the use of visual imagery. Such an interpretation does not seem to recognize the limitations of the tactile sense and the material it has to use. Tactile exploration is a piecemeal affair, and some "tactile" means must exist to integrate the material into a unit to represent the object, if it is to be said that the observer apprehends shape, size, etc., tactually. It would appear that the only mechanism we have for this is visual imagery, which is to say the operation of the visual mechanism itself. It is plausible to think that touch is not a sense mechanism that in and of itself gives any outcome that could be called a spatial appreciation of objects. Examined closely, it would seem to be a sense modality in which the appreciation of mechanical contact with and force against the skin is appreciated. If the psychologist believes that the organism appreciates touched objects as spatial configurations through the operation of tactile mechanisms alone, he appears to be operating on a broad uncritical faith. Be that as it may, it is not the purpose of the present paper to argue the point. No more will be said than is necessary to make clear what is meant by our idea of the possible difference between vision and other senses. The main purpose is rather to report some findings. These were gained from work motivated by an outlook which posits the operation of the visual mechanism as the means whereby the sensing of spatial properties is accomplished. Visual imagery is an example of one output of this mechanism, and in its clearer forms is often involved in the individual's attempt to comprehend object size and form. How lacking clear visual imagery can

---

*Received in the Editorial Office on June 16, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

be in any case, is difficult to say. It would be supposed however that as visual imagery is impoverished, the kind of performance which we can call perceiving size or shape is likewise reduced to the vanishing point.

If the visual mechanism is at work in the structuring of the tactile and kinesthetic information, the behavior resulting from the use of such information should obey the laws of vision. It happens that we know some of the laws of visual perception relating to size, distance, and shape of objects. The operation of visual imagery as such has never been disclosed in terms of specific laws. But whatever they prove to be, they should closely resemble those expressing the operation of perception.

Hence it was the supposition in the present investigation that the perceived size of an object of touch would be dependent upon *where* the object would be when tactually explored. If the object were to be close to the eyes, it would be perceived as one size; if it were to be at a distance it would be perceived as another size. The further away the object, the smaller. This, of course, would have its limits, because tactile exploration, for most practical purposes ends at arm's length. On the other hand, it might be possible to explore the periphery of objects to comprehend their size and shape, with a hand-held stick. This would be as much a matter of kinesthesis as touch, if not wholly the former.

From the sense data of touch and kinesthesis, it was expected that the further away the explored objects, the smaller their perceived size. If so, this would be an entirely unexpected outcome from the conventional premise that size and shape perception from manually explored objects is based on the elaborations of the touch and kinesthetic mechanisms alone.

B. APPARATUS AND PROCEDURE

There are several ways to attempt to determine the relation of perceived size to object distance. Some of these were tried at different times by various students under the direction of the author. None of the experiments were of an extended character, but nevertheless most of them provided unmistakable evidence of the kind needed for checking the general assumption already stated. In most cases, the material consisted of a series of square blocks .75 inch thick, presented in random fashion as to size and distance.

1. *Experiment I*

In this, the observer ran his index finger around the edge of the block twice and indicated the apparent size of the block on a device in which magnitudes could be indicated by the sliding of an element along a bar, to denote separation between it and a stationary element. Thus the separation between the two elements on the bar, as accomplished manually, was the measure. This, likewise, involved visual imagery, except that the slide bar was always at the same distance from the eye. This experiment had among its defects the fact that since the imagery used during "indication" was based upon a fixed distance, this might offset the actual size experience obtained in exploration, and thus reduce or obliterate the size-distance relationship operating in the test situation.

2. *Experiment II*

In another experiment, although the exploration was essentially the same as in *A*, the indication (observer's response) was that of adjusting an Aubert diaphragm to form a square that appeared visually to be the size of the tactually explored square. This experiment was defective, in part, on account of the fewness in number of the explored squares. It was only a little while until this was realized, and the subjects apparently tended merely to "pick out" which of the squares was presented and to use a relatively fixed size for the visual indication, for each block. This is to say that they acted to *identify which* of the few given blocks was presented rather than to *indicate its sensed size.*

3. *Experiment III-a*

In this experiment, the method of constant stimuli was used. A standard 4-inch block was set up at 10 inches from the eye, and one of three other blocks (3½", 4", or 4½") was presented at one or the other of the following distances: 5 inches, 7 inches, 10 inches, 14 inches, or 20 inches. Twenty-five trials for each block for each distance were obtained, for each of two observers. Either one of three responses was allowable. The second block could be "equal to," "smaller than," or '"larger than" the first block in the pair.

4. *Experiment III-b*

At another time, a second experiment, similar to the preceding one, was made, except that only two categories of response were allowed, and only the 3½" and 4½-inch blocks were compared to the standard 4-inch block. In this experiment, the 10-inch distance was also omitted.

5. *Experiment IV*

In a still different experiment the observer was given one of a number of predetermined separations of the forefingers of the two hands, while the fingers touched the edge of the table at which he was seated. He was then instructed to drop his hands to his sides and then bring them up and out over the table until they touched a backboard, and there the forefingers were to be held apart at such a distance as to indicate the separation given them to begin with, by the experimenter. The distance of the backboard was varied from trial to trial. The standard separations given the fingers were: 2 inches, 3 inches, 4 inches, or 5 inches. The distance of the table edge from the eyes was approximately five inches. The backboard was set at either 10 inches, 15 inches, or 20 inches from the eyes. Sixty trials, distributed in five sessions, with 3-minute intervals between, were given. In this experiment, the function tested was primarily a kinesthetic one, but supposedly dependent upon vision (visual imagery, etc.) as in touch.

C. Results

1. *Experiment I*

The judgments here tended to indicate the blocks were perceived as smaller when explored farther from the eye. The results were not statistically reliable. They represented tremendous scatter among observers in the sizes indicated, particularly with those whose behavior was erratic. Of 35 observers, 13 did fairly well. Their results were relatively consistent and in the direction indicated above. The range of sizes as related to distance was not large, however. Actually, we were dissatisfied with the mode of indicating perceived size in this experiment. It had obvious faults, and in light of them, it was not surprising that the results were not markedly and consistently in line with hypothesis.

2. *Experiment II*

There were only six observers and the predominant tendency in response was as indicated earlier. The tendency was to select *which* of the blocks was being presented, and always to record it with approximately the same setting of the diaphragm. This method also had other very obvious faults, although it would be expected that trained observers would have been able to respond on the basis required in the instructions, and might have acted more in line with the expectations of the hypothesis.

3. *Experiment III-a*

Figure 1 indicates the results of this experiment. From it, it will be seen that as distance increases the number of "bigger-than" judgments decrease, and the "smaller-than" judgments increase, and that there is a decided asymmetry in the "equal-to" judgments. Application of the Chi-

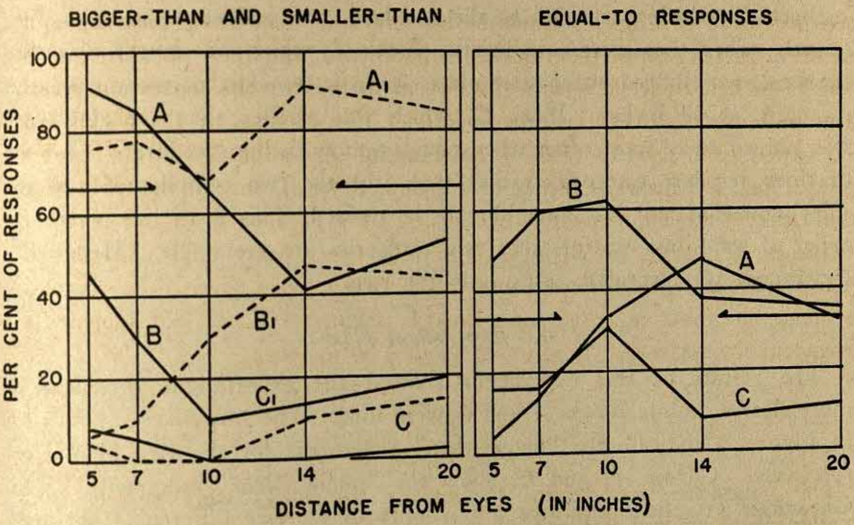


FIGURE 1

Curves (left) indicating judgments of comparison block in relation to a standard 4″ x 4″ block. Curve A is for "bigger-than" judgments, for a block 4½″ x 4½″. Curve C_1 is for "smaller-than" judgments of the same block, under the same conditions (varying distances from the eye for comparison—10″ from eye for standard). Curves B and $B1$, for a 4″ x 4″ block; and Curves C and A_1, for a 3½″ x 3½″ block.

Curves (right) represent the "equal-to" judgments for the same three blocks.

square formula to the data indicated a statistical significance at the 1-per cent level of confidence. Two "*t*" ratios computed one for the difference of a "bigger-than" judgments at 5 and 10 inches, and the other for the "smaller-than" judgments at these distances. The results were also significant at the 1-per cent level of confidence.

In general, the curves support the expectations originally stated, namely that objects appear smaller as they are tactually explored, farther and farther from the eyes. This is based on the finding that the frequency with which blocks are judged larger than the standard block rises as the blocks are brought nearer the eye, the standard remaining at a constant distance. Even smaller blocks are judged larger very frequently when brought nearer the eye than is the standard block. This varies, under the present experimental conditions, for Block B (4 inches)from about 45 per cent of the time to about 10 per cent when it and the 4-inch standard are equidistant. For Block A (4½ inches), the variation was from better than 90 per cent at 5 inches to about 42 per cent when A was at 14 inches. For Block C (3½ inches), which was smaller than the standard, the variation was not great, as would be expected.

Block *B,* which was equal to the standard, was judged equal most frequently, when they were equidistant. Block *A,* which was larger than the standard, was judged equal most often, when it was farther away than the standard, at 14 inches. Block *C,* which was smaller than the standard, was judged equal most often when equidistant with the standard. The use of three response categories rather than just the two complicated the results somewhat and left something to be desired. This is in line with the belief of some investigators that two categories are preferable. Hence the experiment was repeated, using only the two.

4. *Experiment III-b*

The results of this experiment showed the expectations considerably more clearly. Only Blocks *A* and *C* were used. The two curves, *A* and *C,* in Figure 2, indicate the "bigger-than" judgments for the two blocks respectively. Curves A_1 and C_1 show the "smaller-than" judgments. The Chi-square procedure indicated a reliability of the size differences obtained with various distances as being reliable at the 0.1 per cent level of confidence.

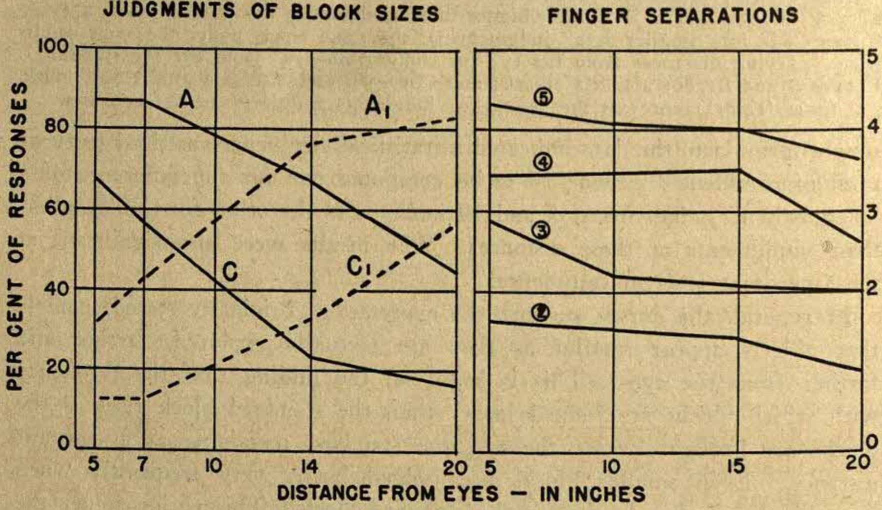


FIGURE 2

Curves (left) showing the judgments of block sizes with reference to a standard block, as related to distances of the blocks from the eyes. Curve *A* indicates judgments of a 4½" x 4½" block as larger than a 4" x 4" standard. Curve A_1, judgments of "smaller-than" under the same conditions. Similarly, Curves *C* and C_1 are for a block 3½" x 3½".

Curves (right) indicate forefinger separations in relation to finger distance from the eyes, for standard "given" separations of 2, 3, 4, and 5 inches respectively. The numbers at the side of graph represent true metric separations. Numbers in circles label the curves for these given separations.

It will be seen that the apparent size of the blocks decreases quite regularly as distance of exploration is increased as determined by frequencies of judging them larger or smaller than a standard block kept at a fixed distance.

5. *Experiment IV*

The curves in the righthand part of Figure 2 describe the results when determining apparently equal separations at various distances from the eye. One might expect that if *size* grows smaller with increase in distance from the eyes, a space interval judged equal to a standard at a fixed distance would have to be greater as distance is increased. This is not exactly the way it works. The observer does not operate that way under the conditions set up in the experiment. He seems, rather, to use the close-up standard as a reference and as he moves his fingers farther away, they are made closer to represent a space which visually (according to the laws of perspective) becomes smaller as viewing distance is increased. This result is a further confirmation of the basic hypothesis that the real frames of reference used in tactile and kinesthetic experience are visual, at least for sighted people. We have some information on block-size, judging by the blind, congenitally and otherwise, but it is not sufficient to report here.

Had we obtained results opposite to those in Figure 2 (right), one still would be inclined to call upon visual imagery for explanation, but the results as they stand are a better confirmation. They show visual imagery as a *guiding influence* in kinesthetic behavior rather than merely as a frame for *interpreting* the behavior as it occurs. This is to suggest that kinesthesis would be a far poorer tool without such guidance than with it. The behavior of the congenitally blind, notwithstanding popular opinion to the contrary, indicates this.

D. SUMMARY AND CONCLUSIONS

In the several experiments performed in the attempt to find whether vision (through its imagery) operated in tactile and kinesthetic experience, it was found that size and space judgments derived from tactile and kinesthetic information were rather gross. Part of this impression stems from the difficulty in obtaining proper indicators for manifesting size and space perceptions. The more appropriate methods eradicate this. The results indicate that visual imagery plays a very significant and indispensable rôle, at least if size and space judgments are to be reasonably stable and reliable. This rôle of visual imagery is evidenced by the lawful change in apparent size with change of distance of exploration from the eye. The results point to the idea that although the blind make extensive use of tactile and kinesthetic information, it must be a sort not describable as the perception of space, but rather the perception of the body moving in the proper way in relation to contact, etc. This is not what we should call true space perception, even though the subjects get around in a spatial world.

Department of Psychology
Michigan State College
East Lansing, Michigan

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 409-415.

CONDITIONS INTENDED TO MANIPULATE ACCOMMODATION AS FACTORS IN PERFORMANCE WITH A VISUAL ACUITY TARGET*

Department of Psychology, Michigan State College

S. HOWARD BARTLEY

A. PROBLEM AND PROCEDURE

The present study was made to determine the effort of accommodation on responses to targets in which visual acuity was a crucial factor. In a sense, the study was an initial step in tracking down the variables that occur in the common psychophysical experiments where response is measured in terms of the percentage correct in a series of trials. In such experiments, there is a marked variation from trial to trial in what the observer (subject) is able to do. In some trials the target is quite clearly visible; in others it is less so, this reduction reaching all the way to complete inability to see the crucial features of the target at all.

Ittelson and Ames (3) asked three pertinent questions in a recent study of the image-producing mechanisms of the eye: (*a*) What is the effect, if any, on apparent distance, of change in the image-producing mechanisms? (*b*) What is the effect, if any, on the image-producing mechanisms of changes in actual distance? (*c*) What is the effect, if any, on the image-producing mechanisms of change of apparent distance?

The findings of these authors indicate that target distance is not a compelling factor in accommodation. For at least some target distances various accommodations and convergences may occur. This is in contradiction to the more conventional views still current (1, 2).

Although the questions of Ittelson and Ames are interrelated, information bearing on any one of them may incidentally arise in studies planned for other purposes. The present investigation is a case of this sort, since two of the variables are those of accommodation and target distance, with the added possibility of apparent target distance differing from real distance.

One of the ways of attempting to control accommodation is to vary the distance of a surface (and its edges) at which the observer's attention is directed. In the present investigation, this was done by using a circular,

---

*Received in the Editorial Office on June 16, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

lighted field on a dark background. In the center of this field was an aperture through which the briefly exposed target was visible. While the target distance was in all cases one meter, the circular field (called the *surround*) was placed at one of three positions nearer to the observer than the target. Three different levels of surround brightness were used, with the assumption that brightness, as well as distance, would manipulate response.

The distances were 67 cm. for Surround *A;* 50 cm. for *B;* and 33.3 cm. for *C*. In all cases, the surrounds subtended a visual angle of 30 degrees. The target subtense was about 1.5 degrees, regulated by the apertures in the surrounds. The surrounds were placed on screens or "shutters" that could be raised or lowered into position in vertical grooves in properly spaced stationary posts. The target and surrounds were viewed through an artificial pupil, 5 mm. in diameter. The surround of high brightness was 0.16 c/ft^2; the moderately bright surround 0.045 c/ft^2; and the low brightness surround, 0.0045 c/ft^2. The brightness of the target was 0.572 c/ft^2.

The observer's response consisted in verbally indicating in which of four positions a hairline was oriented (being either vertical, horizontal, tilted to the left 45 degrees, or to the right 45 degrees).

Three observers were used, and 130 readings were obtained for each condition tested. That is, there were three surround positions, and three levels of target illumination for each. In addition to these, a set of 32 readings was taken when there was no pre-exposure surround either to help determine accommodation *distance,* or to guide *direction* of fixation.

Each of three observers, whose ages ranged from 21 to 26 years, was checked with a Phorometer to determine what lens "correction," if any, would be needed to see best at the surround distances used. These values are indicated in Figure 1.

Effectiveness of performance was measured in terms of percentage of correct responses made under each condition.

The assumptions were as follows: (*a*) That the percentage of correct performance will, in each case, be a measure of the conditions used. (*b*) That distances are more inaccurately perceived and accommodation more grossly adjusted for them when field (surround) illuminations are low than when high. (*c*) That the observer's intrinsic refractive state will be the most effective factor in determining the distance accommodated for at very low illuminations. It will be the sole determinant in darkness. (*d*) At moderate and high levels of illumination, the surround distance itself will tend to determine distance fixated and accommodated for. (*e*) When the surround distance differs from the transient-target distance, the performance

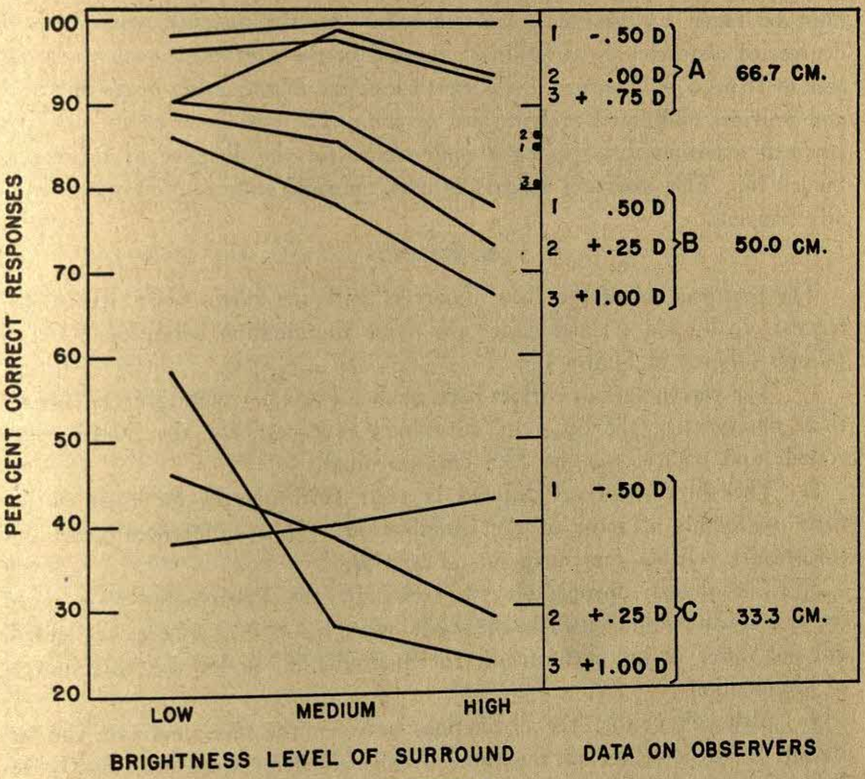


FIGURE 1

Curves showing percentages of correct performance for three observers (1, 2, 3) at each of three brightness levels (low, medium, high), and under three different target-surround situations (*A*, *B*, *C*). The righthand part of chart gives data on refractive corrections needed for best visual acuity for the surround distances in diopters (*D*). The centimeter values indicate distances of surround from observer. The target distance in all cases was one meter.

will be impoverished on account of improper accommodation for target distance. (*f*) If the surround distances used are those for which the observer, on account of his refractive limitations, cannot well accommodate, this factor will be reflected in a drop in percentage of correct performance. (*g*) The fact that the observer can respond correctly sometimes and not at others must be construed as meaning that the factors that enable the performance fluctuate from trial to trial. These have often been conceived of as degrees of *readiness*, but possibly the factor of accommodation, for example, is very loosely related to what is ordinarily understood as readiness.

The various assumptions listed are not all equal in their standing. Item *a* must be taken for granted. Certain others can be assumed with various degrees of assurance from findings reported in the literature, such as is the case in Items *b, c, d, e,* and *f.* Nevertheless, the investigation bears on these and provides additional evidence pro or con. The real issue at stake is the frequent assertion that the eye accommodates for the distance of target attended to. This assertion appears to be a too-rigid statement of what actually happens.

B. RESULTS

The performance of the three observers with the surrounds at three distances: *A, B,* and *C,* and under the three illumination conditions, are as follows (shown in Figure 1):

1. The percentages of correct performance (% c.p.) were greatest for all three observers at the 66.7 cm. surround; next best, for the 50 cm. surround; and poorest for the 33.3 cm. surround.

2. The three observers differed in their performances for each of the three surrounds at most of the illuminations. The differences were not statistically reliable for Surrounds *A* and *B.*

3. The slightly myopic observer who, by the Phorometer test, could see best at all distances used in the study, when a —.50 D sphere was added, did not differ in his performance for Surround *A,* under the three levels of illumination.

4. Although neither the distinctions between the observers nor the behavior of any observer at the three brightness levels were statistically reliable for Surround *A,* the trends with the high illumination were in the expected direction. That is, the "emmetrope's" behavior was between that of the myope and the hyperope.

5. For Surround *B,* there was also no statistically reliable difference between observers in the outcomes, but, here too, the trend was in the expected direction, as for Surround *A.* This trend prevailed not only for the high illumination but also for the other two levels. The myope was consistently better than the other two observers. The "emmetrope" now required a slight plus correction, and the hyperope required a stronger lens than for Surround *A.* The "emmetrope's" performance lay between the myope and the hyperope under all three of the illumination conditions.

6. The data for the high, medium, and low illuminations for Surround *C* were found to be statistically reliable at below the one-per cent level, whereas the data relating to differences between observers was not found to be reliable below the 7-per cent level. This does not take into account the refractive states of the observers. Actually, the observers behaved agair

as would be expected, when this factor was taken into account. For example, the myope's performance was impoverished the least by using Surround *C*, at the highest illumination. In fact, for this observer, it made little or no difference which illumination level was used. This could be taken to mean that he was accommodated for approximately the distance represented by Surround *C*, so that it made little difference whether there was a surround there or not.

7. The other two observers' performances were modified considerably by raising illumination. This might well be expected for the hyperope, if left relatively free to accommodate as he might, that he would tend to accommodate for a distance nearer to that of the target itself (one meter), if not even beyond. The near emmetrope would also, when free, tend to accommodate for some distance greater than the myope. The graphs show his performance to tally with this expectation. Thus, in low illumination, his performance fell between that of the myope and the hyperope. As illumination was raised, he still tended to do this, but the hyperope and the myope exchanged relative positions.

8. The performance of the three observers with no lighted surround is indicated by the three points in the upper righthand side of the graph. With no lighted surround, the observers were not given compelling conditions for fixating at any particular *distance*, nor were they given a stimulus for fixational direction. Trials followed each other in rapid enough succession that little disorientation apparently occurred. It will be noted that the performance values were relatively high.

C. Discussion and Conclusions

We can conclude that as surround distance and target distance became increasingly different, performance is progressively impoverished.

If the target is relatively close to the eyes, namely, one meter or less, the factor of refractive state of the individual observer comes crucially into play. If the surround is nearer the eye than the target, the myope's performance is impoverished the least from the value obtained either with no illuminated surround, or with a surround nearer to the target. The myope is thus at an advantage.

The shift in brightness levels from one very low to one providing for clear observation of the target makes a difference in performance. At the lowest brightness, although the surround can still be seen, its actual distance is possibly somewhat uncertain to the observer. At least, accommodation becomes somewhat inaccurate. The surround does not seem so compelling in determining accommodation for its actual distance from the eye. When surround brightness is raised, performance improves, owing in part to its greater effectiveness in compelling fixation for its distance.

The *target* illumination was, of course, the same for the three levels of surround illumination. The differences in the levels between target and surround might make some difference in the ability to perceive the target. But from reasonable deductions from many studies on performance when target surround relations were manipulated in intensity, the target should be most easily seen when the surround and target are at about the same value. For this reason, the brightest surround should have produced the best results, but it did not for any of the observers. Thus, we can discount the effect of difference between target and surround levels as a major factor in producing the trend of our results.

Therefore, it would appear that, in general, the observers' behavior was consistent with the idea that the surrounds used tended to predetermine the fixation and accommodation distances of the eyes when the surrounds were well illuminated, but progressively less so as illumination decreased. It also appears that lowering of brightness allowed the observer's "refractive states" to become one of the dominant factors in determining performance. If the observer were slightly far-sighted, the close surround, *C,* had more effect in impoverishing performance than if he were either emmetropic, or nearsighted. The low-level surrounds had less effect, and the far-sighted observer, released from much of the surround's control, did better than the others, for he was free to accommodate for some distance near to that of the target.

The results bear on those of Ittelson and Ames, simply showing that the eye does not always fixate for the distance of the "object" attended to, and that this is not always to be attributed to the "refractive state" of the eye.

No control of apparent distance was attempted, and no systematic reports of observers were obtained. In a preliminary experiment, however, a single observer (not one of the present three) was studied. For him, the target distance was 8 feet, and the surround distances were approximately 5 and 3 feet respectively. This observer reported, under several low brightness conditions, that the surround tended to float or appear many feet away from him, while he knew full well that it was actually nearby. He, too, was slightly impoverished in his performance when the surround brightness was raised, when the surround distance was definitely different than that of the target. On the other hand, when target and surround distance were both eight feet, raising the illumination improved performance. In that experiment, the top level of the surround was 2 c/ft^2, and the target was 1.05 c/ft^2. Various other surround levels varying from this down to .00025 c/ft^2 were used. At this latter level, performance was about the same as when the surround was not illuminated at all. From this level on up, there was improvement in performance, at least to .65 c/ft^2.

While the information obtained in the present investigation did not settle the questions of whether, how, and/or how much accommodation shifts from trial to trial, it did show that accommodation was a factor in determining level of performance. It is probable that accommodation does shift from instant to instant as features of the task situation and correlated observer activity run through their cycles from trial to trial. One might expect, however, that with the observer being "ready" for each target exposure as it is made, the variability would be reduced to a minimum.

It would be expected that only performance that is not determined in very transient periods of time (short target exposure) would show much degree of constancy from trial to trial. The shortening of target exposure to provide "minimal stimulation," involves not only less light acting on the eye but a greater statistical range in possible performance, owing to catching the observer in various phases of his oscillatory changes in state. If total stimulation is the crucial factor, one picture of performance variation will result; if distribution of stimulation (as determined in the retinal image by variations of accommodation) is crucial, then another trial-to-trial pattern of performance will eventuate.

REFERENCES

1. DIMMICK, F. L. *In* Boring, Langfeld, and Weld. *Foundations of Psychology.* New York: Wiley, 1948.
2. GRANT, V. W. Accommodation and convergence in visual space perception. *J. Exp. Psychol.,* 1942, **31**, 89-104.
3. ITTELSON, W. H., & AMES, A., JR. Accommodation, convergence, and their relation to apparent distance. *J. of Psychol.,* 1950, **30**, 43-62.

Department of Psychology
Michigan State College
East Lansing, Michigan

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 417-422.

A PRELIMINARY ATTEMPT TO IDENTIFY THE FACTORS IN STUDENT-INSTRUCTOR EVALUATION*

Department of Psychology, Miami University

C. W. Crannell

A. Introduction

The purpose of this study was to attempt a tentative identification of the factors involved in appraisal by college students of the performance of their instructors. By discovering such factors (in the present case it may be better to speak of "clusters of components") it may become possible to construct an instructor rating device which would cover as many aspects of instructor performance as a student may properly be expected to appraise, and to do this with as few items as possible, so as to minimize the expenditure of time in making the ratings and in their subsequent analysis.

In the present study an experimental form of the Miami University Instructor Rating Sheet was prepared which contained seven new items in addition to the original 14. The reader is referred to a report by the writer (5) on the original version, and also to a number of studies by Bendig (1, 2, 3) and Russell and Bendig (7). These authors have shown the value of increasing the discriminatory power of the items by rewording the identifying statements so as to create a better spread of responses than was found with the original rating sheet. They have also demonstrated that there is only a slight relationship between grades in courses and the students' appraisals of their instruction (7).

B. Procedure

The seven new items added to the original rating sheet were selected after examining critical appraisals submitted by many members of the local teaching staff. Of great help in preparing these items was the *University of Minnesota Student Report on Classroom Teaching, Form A,* prepared by Clark (4), which contains 23 well constructed evaluation items and several questions about the student's own status in college. The experimental form was mimeographed and used in certain classes at the close of the second semester, 1950, at a time when the regular printed form

---

*Received in the Editorial Office on June 16, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

was used in all other classes in the College of Arts and Sciences. Of 400 copies of the experimental form, over 300 were returned filled out, and, after elimination of any rating sheet with one or more items unanswered, a sample of exactly 300 sheets was made up. This sample was taken from 13 different courses taught by nine different instructors in Chemistry, English, Mathematics, Psychology, and Speech. In terms of course level, 125 were from freshman courses, 91 from sophomore courses, and 84 from upper division courses.

A tabulation was made of each item in terms of the five categories of response. On the basis of these tabulations the responses were dichotomized into high and low groups, as nearly even in number as the data would permit. This information was then punched on 300 Keysort cards. All sorts and counts were made, and tetrachoric correlations computed, using the Hayes charts (6).

When the correlation matrix was completed[1] a cluster analysis following Tryon's method (8) was performed. This analysis was carried only through the stages of allocation of variables to clusters (8, pp. 42-43) and trial correlation profiles of variables with mean cluster correlations (8, p. 53). Several possible alternative mean groupings were tested in order to determine whether or not the congruent profiles would be altered radically if the analysis assigned certain ambiguous variables to different clusters.

A list of the 21 items of the experimental form follows:

a. Organization of course material.
b. Content of classroom presentation. (In Form *B*, the word, "content," has been changed to "interest" so as to be more consistent with the meaning of the five responses.)
c. Attitude toward questions.
d. Attitude toward students (friendliness).
e. Attitude toward student effort.
f. Control of class.
g. Length of assignments.
h. Nature of examinations (fairness).
i. Grading system (impartiality).
j. Sense of humor.
k. Orientation of subject matter (usefulness).
l. Mannerisms.
m. Amount being learned.
n. Preparation for class meetings.
o. Demand for original thinking.
p. Relation to other areas of knowledge.

---

[1] A copy of this matrix will be provided by the writer upon request.

q. Tolerance of student opinion.

r. Voice.

s. Would you recommend this course to a good friend whose interests were like yours?

t. Contribution of textbook.

u. Overall rating of the course.

C. RESULTS

The highest correlation obtained among the 21 items of the experimental form was .80, between Items *s* and *u*. The highest negative correlation was —.25, between Items *o* and *t*. The highest average intercorrelation with all other 20 items was .419, obtained for Item *m*. The lowest average intercorrelation was .006, obtained for Item *t*. Of the 210 correlations, 103 were below .30, 59 were below .20, 33 were not significant at the 5 per cent level, and three were significantly negative.

The cluster analysis revealed the presence of five clusters, although analysis into three main clusters was quite reasonable. Cluster C-1 contained Items *b, k, m, s,* and *u*. Cluster C-2A contained Items *c, d, e, j,* and *l*. Cluster C-2B contained Items *o* and *p*. Item *q* could be made to fit equally well in either C-2A or C-2B. When tested against mean correlation profiles for an arrangement in only three clusters, the last three items fitted well enough to suggest that most of the components involved in these items can be accounted for in terms of a single C-2 cluster. Cluster C-3A contained Items *a, f, n,* and *r*. Cluster C-3B contained the Items *g, h,* and *i*. This group also fitted well in a three-cluster arrangement with the other C-3 items. Item *t* had such low correlations as not to fit too well in any cluster, but its mean profile was most congruent with Cluster 3.

This study has made it possible to assign names to three tentative groups of components involved in student appraisal of their instructors. Cluster C-1 may be termed "Course result: what the student thinks he is getting out of the course." It is worth noting that the intercorrelations among these items were in general the highest within the matrix. Cluster C-2 may be termed "Personal interaction: the relationship the student feels he has with the instructor as a person." Cluster C-3 may be termed "Instructor effort: what the student thinks the instructor is putting into the course."

Whether or not these clusters would hold up as "factors" in an orthometric analysis is a matter for future study with larger groups of raters. It is planned to perform such a treatment using the revised rating sheet, to be described below. In any event, these tentative clusters do appear to make "psychological sense," and the study demonstrates that there is a type of consistency in student-instructor appraisal which is more than a mere halo effect, is indicative of discriminative behavior on the part of student raters, and is amenable to quantitative evaluation.

Analysis of the data reported above led to the development of the Miami University Instructor Rating Sheet, Form *B* (Figure 1). This new version retains nine of the original 14 items: 1, 2, 3, 4, 5, 6, 12, 13, and 14, and includes five new items from the experimental form: 7, 8, 9, 10, and 11.

PLACE A CHECK MARK IN ONE SPACE FOR EACH OF THE FOLLOWING ITEMS. BE SURE NOT TO OMIT ANY ITEM.

1. Organization of Course Material	Clear and definite continuity from day to day	A well systematized presentation of the material	Well organized most of the time	Part well, part poorly organized	Often indefinite and confusing
2. Interest of Classroom Presentation.	Always very interesting	Interesting most of the time	Interesting more often than not	Often somewhat uninteresting	Often very dull
3. Attitude toward Questions by Students	Unusually skilled in drawing out questions	Always encourages questions	Shows reasonable interest in students' questions	Sometimes seems uninterested in questions	Avoids many reasonable questions
4. Attitude toward Student Effort	Exceptionally appreciative of effort	Encourages students to work hard	Usually shows appreciation of effort	Rarely shows much appreciation of effort	Indifferent to student effort
5. Nature of Examinations	The fairest examinations I have taken	Usually very good in wording and content	Examinations are fair and reasonable	Sometimes not as fair as they could be	Clearly unfair to many students
6. Grading System	Perfectly fair and impartial	Very fair and impartial	Reasonable for most students	Unfair to some students	Grades seldom reflect student achievement
7. Amount Being Learned	Have learned more than from any other instructor	Have learned more than from most instructors	Have learned an average amount	Have learned less than I feel I should	Have learned less than from most instructors
8. Preparation for Class Meetings	One of the best prepared instructors	Better prepared than most instructors	As well prepared as the average	Less well prepared than the average	Not as well prepared as most instructors
9. Demand for Originality	Emphasis on original thinking is always maintained	Much original thinking is demanded	Moderate amount required	Very little demand for original thinking	No original thinking required
10. Relation to Other Areas of Knowledge	Effectively relates his presentation to other areas of knowledge	Often shows how his subject is related to other areas	Presents relationships only to a limited extent	Most of the time keeps strictly within his own area	Teaches a narrow overspecialized course
11. Voice and Diction	Always uses excellent enunciation	Very clear and audible	About average among my teachers	Sometimes has rather unclear enunciation	Requires considerable effort to understand
12. Mannerisms	One of the most pleasing instructors in actions and appearance	Very pleasant actions and manner	No unusual actions	Occasional display of mannerisms	Certain clearly annoying mannerisms

(If you checked one of the last two spaces in 12, please indicate briefly what you mean)

....

....

13. Contribution of Text	One of the most valuable parts of course	A good, sound textbook	Textbook is fairly useful	Textbook not very adequate	Of little value in contributing to course
14. Overall Rating of Course	An outstanding course in the college	One of the better courses I have taken	A reasonably good course	Just to be endured as one more course	A rather poor course

15. Comments. (*Note*: Experience with this form has shown student comments to be of great value). Add anything here which you feel is not included in the scale, but necessary to your evaluation of the instructor..

....

....

....

FIGURE 1

Before printing the new form a mimeographed edition was tried out in the latest summer sessions, and a response frequency analysis of 189 returns from 20 different classes in seven different departments indicates that the revision should have less skew in the responses than was found with the original form. About four thousand copies are currently being distributed among members of the faculty of the College of Arts and Sciences, with a view to collecting data for the next analysis.[2]

D. SUMMARY

An analysis of an experimental rating sheet containing 21 items, administered in several different classes to 300 students, led to the tentative identification of three clusters of components, or "factors," involved in student appraisal of their instructors. These clusters have been named "course result," "personal interaction," and "instructor effort." A revised 14-item form of the rating sheet was constructed, with which it is planned to do a more elaborate analysis involving much larger samples.

This type of research certainly should be undertaken wherever any form of student-instructor appraisal is employed. The writer is not convinced that, as of this writing, there exists any such rating procedure so well quantified and analyzed as to yield data which may be interpreted without ambiguity. It is therefore no matter of surprise that many of our faculty colleagues look askance at the efforts to further student-instructor evaluation, when the procedures we are able to offer them have little or no basis in quantitative research. The complexity of human behavior, nowhere greater than in the student-instructor relationship, demands that investigators pursue a continuing program of defining the concepts and refining the methods. To rest on the weak foundation we now have is worse than to abandon the whole task.

REFERENCES

1. BENDIG, A. W. A preliminary study of the effect of academic level, sex, and course variables on student rating of psychology instructors. *J. of Psychol.*, 1952, **34**, 21-26.
2. ———. The use of student rating scales in the evaluation of instructors in introductory psychology. *J. Educ. Psychol.*, 1952, **43**, 167-175.
3. ———. A statistical report on a revision of the Miami Instructor Rating Sheet. *J. Educ. Psychol.*, 1952, **43**, 423-429.
4. CLARK, K. E. The University of Minnesota Report on Classroom Teaching. Form *A*. Department of Psychology, University of Minnesota, 1948.

---

[2] The writer is indebted to the Miami University Alumni Loyalty Fund for a grant which makes possible continuation of this research.

5. CRANNELL, C. W. An experiment in the rating of instructors by their students. *College and Univ.*, 1948, **23**, 5-11.
6. HAYES, S. P., JR. Diagrams for computing tetrachoric correlation coefficients from percentage differences. *Psychometrika,* 1946, **11**, 163-172.
7. RUSSELL, H. E., & BENDIG, A. W. The relation of student achievement indices of teaching effectiveness and student rating of introductory psychology instructors. *Amer. Psychol.,* 1942, **7**, 246. (Abstract; MS seen by the writer.)
8. TRYON, R. C. Cluster Analysis. Ann Arbor: Edwards, 1939. Pp. 122.

Department of Psychology
Miami University
Oxford, Ohio

Published as a separate and in *The Journal of Psychology*, 1953, **36**, 423-425.

SEASON OF BIRTH AS RELATED TO SEASONAL PREFERENCE AND PERSONALITY TRAITS*

Department of Psychology, Howard University

CHARLES E. MIDDLETON AND F. C. SUMNER

A. PROBLEM AND PROCEDURE

The present study of the relation of natal season to seasonal preference and to certain personality traits was suggested by the recent controversy of J. de Sauvage Nolting and Wendel J. Pile as to the relation of natal month to incidence of schizophrenia (dementia praecox).

Nolting's study in Holland disclosed that the number of dementia praecox patients born in winter is significantly greater than that of such patients born in summer and that the month of March has the highest frequency as birth month of such patients while June has the lowest. Nolting would attribute his findings to various effects of weather upon purely chemical processes of conception and intrauterine life, upon psychological aspects of early parent-child relationships, and upon the varying needs for survival as determined by seasonal change.

Pile's data gathered from mental hospitals, white and Negro, of the State of Virginia furnish no support to Nolting's findings. For 1,469 dementia praecox patients from mental hospitals in both eastern and western Virginia the seasonal distribution of birth dates is found uniform both as to region and race.

Questions raised in the mind of the present authors by the contradictory results obtained by Nolting and Pile are: (*a*) whether diagnostic differences may account for the discrepancy in results; (b) whether, somewhat aside from the point immediately at issue, there may exist in "normals" any association of natal season either with seasonal preference or with certain personality traits (for example, self-confidence and sociability). In an exploratory way answer to the second of the two questions is sought in the present study.

Subjects were 192 normal, predominantly young women residing in the area of New York City and ranging in age from 16 to 45 (mean age 17.8). The Bernreuter Personality Inventory was administered and on a separate

---

*Received in the Editorial Office on June 17, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

sheet the subjects were asked for certain personal data including their birth date and the ranking of the four seasons as to their preference. Of the 192 subjects, 45 were born in spring, 48 in summer, 46 in autumn, and 53 in winter.

B. Results

1. The group expressed a seasonal preference in the following decreasing order: Spring, Summer, Autumn, Winter. The spacing of the group's seasonal preferences in terms of the sums of the seasonal rankings may be seen in Table 1. Seventy-seven members of the group preferred spring; 66 summer; 41 autumn; and 8 winter. Winter is by far and large the least preferred of the seasons.

2. The correlation between most preferred season and the season of birth is found to be —.003 or the equivalent of zero.

3. The mean F1C (self-confidence) score of those born in one season did not differ significantly from the mean F1C score of those born in another season with the exception namely, that the mean F1C score of those born in winter was significantly higher at the 5 per cent level of confidence ($t = 2.01$) than the mean F1C score of those born in autumn.

TABLE 1

Scale of Seasonal Preferences in Terms of the Sums of Seasonal Rankings

350	(361)	Spring	1st Choice
400			
	(445)	Summer	2nd Choice
450			
	(471)	Autumn	3rd Choice
500			
550			
600			
650	(655)	Winter	4th Choice
700			

4. In no instance did the mean F1C score of those most preferring one season differ significantly from the mean F1C score of those most preferring another season.

5. In no instance did the mean F2S (sociability) score of those born in one season differ significantly from the mean F2S score of those born in another season.

6. The mean F2S score of those most preferring Spring was found significantly higher at the 1 per cent level of confidence ($t = 4.17$) than the mean F2S score of those most preferring winter; the mean F2S score of those most preferring summer is significantly higher at the 1 per cent level

of confidence ($t = 3.00$) than the mean F2S score of those most preferring winter; the mean F2S score of those most preferring autumn is significantly higher at the 1 per cent level of confidence ($t = 3.10$) than the mean F2S score of those most preferring winter.

C. CONCLUSIONS

Summarized, the conclusions of this study of a group of young women are to the effect that (*a*) seasonal preferences run from a high for spring, with summer a close second, and autumn a close third, to a very low for winter; (*b*) no relationship exists between preferred season and season of birth; (*c*) season of birth makes no difference in self-confidence save that those born in winter are significantly more self-confident than those born in autumn; (*d*) no difference in self-confidence exists between those preferring one season most and those preferring another season most; (*e*) no difference in sociability exists between those born in one season and those born in another season; (*f*) those most preferring spring, summer, or autumn are decidedly more sociable than those most preferring winter.

As to the Nolting vs. Pile controversy the present findings have little to offer in the way of a settlement. Assuming that sociability is a personality trait not associated with schizoid tendency and that self-confidence is a trait associated with schizoid tendency, one might venture to say (*a*) that season of birth has nothing to do with either of these personality traits save a barely significant tendency for those born in winter to be more self-confident, hence schizoid, than those born in autumn; (*b*) that sociability, regardless of natal season, is highly associated with preference for spring, summer, and autumn but not for winter.

REFERENCES

1. NOLTING, J. DE S. Correlation between dementia praecox and month of birth. *Nedrl. tidjschr. v. geneesk,* 1934, **78**, 528-530.
2. ———. Further studies on correlation between schizophrenia and month of birth. *Nedrl. tidjschr. v. geneesk,* 1939, **83**, 3219-3222.
3. PILE, W. J. A study on the correlation between dementia praecox and the month of birth. *Virginia Med. Mo.,* 1951, **78**, 438-440.

Department of Psychology
Howard University
Washington 1, D. C.

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 427-433.

STUDENT ACHIEVEMENT IN INTRODUCTORY PSYCHOLOGY AND STUDENT RATINGS OF THE COMPETENCE AND EMPATHY OF THEIR INSTRUCTORS*

Department of Psychology, University of Pittsburgh

A. W. BENDIG

A. THE PROBLEM

In a recent article (5) the results of a factor analysis of the mean ratings of a group of instructors in introductory psychology were reported. Three factors were found to account for the intercorrelations of the first 10 scales on the Purdue Rating Scale for Instruction: (*a*) a general "halo" factor pervading nine of the 10 scales, (*b*) a group factor with loadings on four scales that was called "Instructional Competence," and (*c*) a group factor correlating with four other scales that was labelled "Instructor Empathy." Multiple regression equations were given for estimating factor scores for each of the last two orthogonal factors from the mean ratings on three rating scales. Such estimates were shown to have a factorial validity of 0.86.

One of the areas of research suggested was an investigation of the relation between these two teaching factors and the achievement of students in the course. Previous research (4) has indicated a low relationship between student achievement and their rating of the instructor and of the course among the students of any single instructor, but a high positive relationship between mean achievement of several instructors' students and mean ratings of the instructors and the course. In this previous study (4) the measure of student evaluation of the instructor was a simple summing of ratings on the first 10 scales of the P.R.S.I.; a measure of instructor effectiveness that the later study (5) showed to be factorially complex. Whether similar relationships would be found between achievement and the orthogonal components in the ratings formed the question investigated in the present research.

B. PROCEDURE

Ratings were collected from the students of five introductory psychology instructors who taught all of the daytime sections during the Fall, 1951, semester. Student responses were gathered near the end of the semester by

---

*Received in the Editorial Office on June 19, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

the present author, rather than the five instructors involved. The students were requested to sign their names to the ratings on the Purdue Rating Scale for Instruction since this was necessary for research on the adequacy of the scales. They were assured that their instructor would never see individual ratings, but if they feared that their ratings might affect their course grade they had the option of omitting their name. Parenthetically, it might be noted that no unsigned P.R.S.I. answer sheets were turned in.

Information had been collected at the beginning of the semester about characteristics of each student, including age, sex, academic level, major, etc., and these data were available for this study. In addition, course achievement information as represented by letter grades was available on each student at the conclusion of the semester. These letter grades have been shown to have a reliability of 0.80 (3).

One difficulty in the analysis was the lack of homogeneity between instructors in the academic level of their students. Some instructors had few juniors or seniors in their classes, others had many. Two recent researches (1, 8, p. 30) offer evidence that juniors and seniors are more critical in evaluating instructors than are freshmen and sophomores, consequently it was felt that this variable should be controlled. Because of the above-mentioned data limitations the solution decided upon was to limit the analysis to the ratings of a homogeneous group; the sophomores. The 121 sophomore men and women taught by these five instructors represent the largest single academic level group enrolling in introductory psychology at this university, comprising approximately 49 per cent of the total daytime enrollment (2).

The student ratings were quantified by applying the regression weights given in the previous publication (5) to predict factor scores. It must be noted that the regression equations used here were originally devised to be used with *mean* ratings on certain scales of the P.R.S.I. Here the equations were used with the ratings of a single student as the best weights for predicting factor scores. It might be pointed out that whether the equations are applied to single student ratings or to mean ratings that the mean factor scores for any instructor would be the same. Since the equations were devised to yield independent measures of instructoral competence when applied to mean ratings rather than single ratings, we might expect a low intercorrelation between the mean factor scores of these five instructors, but some appreciable correlation between the factors within each instructor. This is confirmed by a correlation of —0.10 between mean instructor ratings on Factors II and III and an average correlation of 0.54 between factor scores within the instructors' classes.

The single achievement letter grade for each student was quantified by giving an *A* grade a numerical value of 4, and *B* grade a value of 3, etc. This is the quantification method usually used in computing academic quality points averages by most colleges and universities and was used in the study of letter grade reliability previously referred to (3).

C. RESULTS

Separate analyses of variance were first computed on the separate Factor II and Factor III scores, considering differences between instructors and differences between the sex of the students as independent variables. In addition, analyses of covariance between achievement grades in the course and the two factor scores were also computed (7, pp. 235-240). Table 1 gives the sums of squares for Factor II and Factor III scores and for the quantified letter grades, plus the sums of cross products for the covariance of grades and factor scores. The mean squares used to test the significance of the independent variables can be easily derived from the data given in Table 1. Differences between instructors on Factor II were significant at the .05 confidence point and instructor differences on Factor III at the .01 point. The interaction of instructors and student sex was significant at the .01 point for Factor II, but was not significant for Factor III. When differences between the mean level of achievement of the students of the five instructors were statistically eliminated by the covariance analysis, the significant differences between instructors on Factor II disappeared, while the instructor differences on Factor III remained significant at the .01 point. This was a result of the high negative correlation between the mean achievement of the instructors' students and mean Factor II scores (—0.80) and the zero correlation between mean grade and mean Factor III scores. The significant instructor-student sex interaction for Factor II remained significant when the covariance adjustment was made.

An analysis of variance of both Factor II and Factor III scores was calculated and can be found in Table 2. An analysis of the main effects shows differences between instructors significant at the .01 point, differences between factors significant at the .05 point, and no difference between men and women students when both factors are combined. The interaction of factors and instructors was significant at the .01 point, and, as might be predicted from the analyses of variance in Table 1, the triple interaction of factors, instructors, and student sex variables was significant at the .05 point. None of the other interactions showed significance.

Chi-square tests of homogeneity of variance (6, pp. 195-198) within the

TABLE 1

ANALYSES OF VARIANCE AND OF COVARIANCE OF LETTER GRADES AND FACTORS II AND III FOR STUDENTS IN INTRODUCTORY PSYCHOLOGY

		Original sums of squares				Adjusted sums of squares		
	Source of variation	*df*	Grade	Cross products	Factor	*df*	Factor	*r*
	Total	120	131.32	19.32	229.39	119	296.55	0.10
	Instructor	4	2.39	— 5.84	22.08*	3	7.81	—0.80
Factor II	Sex	1	1.22	— 0.40	0.14	0	0.01	—1.00
	I x S	4	2.26	1.55	43.52**	3	42.46**	0.16
	Within	111	125.45	24.01	233.65	110	229.05	0.14
	Total	120	131.32	21.68	479.32	119	475.74	0.09
	Instructor	4	2.39	0.05	85.09**	3	85.09**	0.00
Factor III	Sex	1	1.22	— 0.78	0.51	0	0.01	—1.00
	I x S	4	2.26	1.37	27.03	3	26.20	0.18
	Within	111	125.45	21.04	366.69	110	363.16	0.10

*Derived mean square significant at the 0.05 confidence point.
**Derived mean square significant at the 0.01 confidence point.

TABLE 2
ANALYSIS OF VARIANCE OF FACTOR SCORES

Source of variation	Sum of squares	*df*	Mean square	*F*
Total	784.00	241		
Between Students	551.79	120		
Instructors	62.60	4	15.65	3.79**
Sex	0.59	1	0.59	——
Interaction: I x S	29.74	4	7.44	1.80
Residual (error)	458.86	111	4.13	
Within Students	232.21	121		
Factors	5.30	1	5.30	4.17*
Interaction: F x I	69.74	4	17.44	13.73**
F x S	0.06	1	0.06	——
F x I x S	15.62	4	3.90	3.07*
Residual (error)	141.49	111	1.27	

*Significant at the .05 point.
**Significant at the .01 point.

various subgroups gave no evidence of variance heterogeneity for either of the factor scores or for grades.

D. DISCUSSION

We have previously suggested (5) that group Factor II found in a factor analysis of the first 10 scales of the Purdue Rating Scale for Instruction could be named "Instructional Competence" and Factor III called "Instructor Empathy." These appellations were defended on the basis of (*a*) a rational analysis of the particular scales heavily loaded with each factor, and (*b*) the similarity between these factors and those found in other factor analyses of teacher rating scales. The present study was designed as an investigation of the relationship of these two P.R.S.I. factors to student sex and course achievement and of the interrelationship of the factors.

Table 1 indicates that both factors discriminate individual differences between instructors and, taking the instructors as a group, both men and women students rate instructors similarly on both factors. However, there is a definite sex difference in the ratings of men and women students on individual instructors' Competence. One instructor may be rated highly competent by his male students and low in the same factor by the women, whereas the reverse may be true for another instructor. Moreover, this phenomenon is independent of any achievement differences between men and women. This suggests that the empathy of an instructor is evaluated the same by both men and women students, but that they disagree on their judgments of competent teaching behavior of instructors. Future research might be directed toward discovering what specific instructoral behaviors are considered to be desirable by women students as compared with behavior judged desirable by men.

The high negative correlation between mean student achievement and mean rating on Instructional Competence is somewhat surprising. An instructor whose students achieve at a high level tends to be rated as being less competent as a teacher than an instructor whose students perform less well on objective achievement tests. Students of high overall ability may be more aware of inadequacies in the teaching ability of their instructors and to judge them more critically. However, student achievement is unrelated to the judgments of instructors' Empathy; mean grade correlating zero with mean ratings of the instructors' attitudes toward their students. The results in Table 2 confirm the ability of the factors to discriminate differences between instructors. Also note that significant patterns of factor and student sex differences existed among the five instructors in the sample.

The present results confirm the hypothesis that the previously reported (4) strong positive correlation between students' average achievement and summed ratings on P.R.S.I. scales was a function of the factorial complexity of the scales. The general "halo" factor, Factor I, found in nine of these scales (5) and not used in the present analysis is probably highly and positively correlated with student achievement. Since Factor I accounts for a larger percentage of the total variance of the ratings (30 per cent) than either of the other two factors, its exclusion from the ratings that were analyzed here would seem to account for the apparent discrepancies between the results of the two studies.

Considering all of the analyses the following generalizations can be made:

1. Sophomore students in introductory psychology can reliably discriminate individual differences in their instructors' teaching competence and empathy.
2. Students' mean judgments of instructional competence are highly and negatively related to the average course achievement of the students, but there is no relationship between student achievement and ratings of an instructor's empathetic attitude toward his students.
3. No general student sex differences are present in ratings of either instructors' competence or empathy, but a particular instructor's competence may be rated quite differently by his men and women students.

E. SUMMARY

The ratings by 121 sophomore students of five introductory psychology instructors were used to estimate the instructional competence and empathy of the instructors. Analyses of variance indicated significant differences between instructors on both traits, no overall difference between men and women student ratings for either trait, but a significant interaction term between instructors and student sex variables for the trait of competence. Covariance analysis showed a high negative correlation between mean student achievement and mean ratings of instructional competence, but none of the other rating variables were significantly related to student achievement.

REFERENCES

1. BENDIG, A. W. A preliminary study of the effect of academic level, sex, and course variables on student rating of psychology instructors. *J. of Psychol.*, 1952, **34**, 21-26.
2. ———. A population survey of students enrolling in introductory psychology at the University of Pittsburgh. (Unpublished study, 1952.)
3. ———. The reliability of letter grades. *Educ. Psychol. Meas.*, 1953, **13**, 311-321.
4. ———. The effect of level of achievement upon students' instructor and course ratings in introductory psychology. *Educ. Psychol. Meas.*, 1953, (in press).
5. ———. A factor analysis of student ratings of introductory psychology instructors on the Purdue Rating Scale for Instruction. *J. Educ. Psychol.*, (in press).
6. EDWARDS, A. L. Experimental Design in Psychological Research. New York: Rinehart, 1950. Pp. 446.
7. ENGELHART, M. D. Suggestions with respect to experimentation under school conditions. *J. Exper. Educ.*, 1946, **14**, 225-244.
8. SMIT, J. A. A study of critical requirements for instructors of general psychology courses. Unpublished Ph.D. dissertation, Univ. of Pittsburgh, 1951.

Department of Psychology
University of Pittsburgh
Pittsburgh 13, Pennsylvania

Published as a separate and in *The Journal of Psychology,* 1953, **36**, 435-448.

THE PSYCHOPHYSIOLOGICAL EFFECTS OF TEA*

Department of Psychology, Brown University

WALTER C. STANLEY AND HAROLD SCHLOSBERG[1]

A. THE PROBLEM

Tea as a beverage was first mentioned in the literature as having been used by man as early as 2737 B.C. Although infusions made from all kinds of leaves, seeds, berries, etc., have been imbibed sporadically throughout the centuries, it is most interesting that tea has had a continuing universal and widespread appeal down to the present time. To what attributes does tea owe this unusual continued acceptance?

One apparent explanation for this acceptance of tea has been the presence of caffeine—a substance known to produce definite physiological effects. Among the studies on the effects of caffeine have been some concerned with various psychophysiological functions such as speed of reaction and mathematical work. Such studies, often using considerably more caffeine than is ordinarily contained in a cup of tea, usually report a small stimulating effect, becoming apparent a half hour or more after ingestion of the caffeine (3). Certainly this delayed effect does not explain what tea drinkers experience. They report a substantial "lift" which is both immediate and sustained. Apparently the many other factors in a cup of tea must contribute to these effects. There are a variety of substances other than caffeine in this beverage which possess physiological activity.

In addition to the physiological effects due to the various factors present in a cup of tea, consideration should also be given to the psychological effects derived from the drinking of the beverage in accounting for the effects produced by a cup of tea. Taste is extremely important to the tea drinker. It is probably an acquired taste, based ultimately on the effects of the constituents in tea. In this sense, the taste of tea is a learned reward in contrast to the taste of sugar which is probably a primary or innate reward. Undoubtedly, the taste of tea is a composite learned reward, since there are other values associated with drinking tea, such as pleasant conversation, rest from work, etc.

---

*Received in the Editorial Office on June 20, 1953, and published immediately at Provincetown, Massachusetts. Copyright by The Journal Press.

[1] We wish to express our gratitude to Thomas J. Lipton, Inc., for financial assistance. We also wish to thank Willard F. Greenwald and Frank P. Panzarella for suggesting the project and for their advice on many aspects of the problem.

The present experiment was directed to measuring the immediate (including the associated) effects of tea as well as the delayed effects. The basic experimental design called for a variety of tests of psychophysiological functions, to be administered several times during a session. Some fatiguing tasks were inserted between the tests. The basic variable, a pause with tea, was placed roughly midway in the session. To measure the effects of a cup of tea, it was necessary to have control sessions, matching the experimental session in all details except that no tea was served during the pause. A preliminary experiment was carried out on six subjects, in which many tests were tried. On the basis of this study, a final battery of tests was selected which will be described below.

B. PROCEDURE

1. *Subjects*

Twenty-two housewives who were habitual tea drinkers (one or more cups a day) served as subjects. Their average age was 40. Most of the subjects were non-smokers; all habitually drank tea without smoking.

2. *Control for Variables Other Than Tea*

The breakdown of subjects given in Table 1 indicates the degree to which variables other than tea were balanced out or controlled. For the nine subjects habitually drinking tea in the afternoon, use of sugar and sessions during which tea was served were almost balanced (one exception). The results of these subjects serve as one check on the lack of balance of these two variables present with respect to the 14 mealtime tea drinkers. It will be shown that this imbalance does not account for test data which are statistically significant. Analogously, a check on the effect of with-

TABLE 1
BREAKDOWN OF SUBJECTS BY CATEGORIES

Habitual hour for drinking tea	Use of sugar	Days on which subject had tea	Number of subjects
Afternoon	Sugar	T—T*	3
Afternoon	Sugar	–TT–**	2
Afternoon	No sugar	T—T	2
Afternoon	No sugar	–TT–	2
Mealtime	Sugar	T—T	5
Mealtime	Sugar	–TT–	2
Mealtime	No sugar	T—T	4
Mealtime	No sugar	–TT–	2

*Tea on Days 1 and 4.
**Tea on Days 2 and 3.

holding tea from habitual tea drinkers was provided by having about half the subjects habitual afternoon drinkers and about half habitual mealtime drinkers. Since all but four subjects were run in the afternoon, any such "deprivation" effect should show up more prominently in the habitual afternoon subjects than in the habitual mealtime subjects. For this breakdown, the four subjects run in the morning were classified as habitual mealtime drinkers, since they received tea during a non-habitual hour.

3. *Schedule of Sessions*

Two experimenters[2] tested the subjects in pairs during five successive days, Monday through Friday. The first day served as an indoctrination session. After the description of the experiment as "having to do with variability on tests," psychogalvanometer electrodes[3] were taped to the soles of the subjects' feet. Then the subjects filled out questionnaires concerning their beverage and smoking habits. About an hour and a half of practice was given on the various tests to be used. One cup of tea was served about halfway through the session.

The schedule of testing (Table 2) remained essentially the same throughout each subject's four experimental sessions. The two exceptions were: (*a*) tea was served during pauses on only two of the days; (*b*) the attention test was given about 20 per cent faster on Days 3 and 4 than on Days 1 and 2, in an effort to compensate for practice effects.

TABLE 2
SCHEDULE OF TESTING WITHIN A SESSION

Test Period 1	15 minutes	2:00-2:15 P.M.*
Task Period 1	15 minutes	2:15-2:30
Test Period 2	15 minutes	2:30-2:45
Pause (with or without tea)	15 minutes	2:45-3:00
Test Period 3	15 minutes	3:00-3:15
Task Period 2	15 minutes	3:15-3:30
Test Period 4	15 minutes	3:30-3:45
Task Period 3	15 minutes	3:45-4:00
Test Period 5	15 minutes	4:00-4:15

*Morning sessions began at 9:00.

During each test period, one experimenter administered the steadiness, simple reaction time, and complex reaction time test to one subject, while the other experimenter obtained subjective ratings of fatigue and tension and administered the strength of grip and attention tests to the other subject. The subjects then changed places and each experimenter repeated her measurements on her second subject to complete a test period. Following Test Periods 1, 3, and 4, both subjects were given 15 minutes to work on long division problems (Task Periods 1, 2, and 3, respectively). After each task period, the subjects again changed places so that their order of taking the tests within each period was held constant. The pause came after the second test period.

---

[2] Mrs. Roberta Nachbar and Miss Frances Clayton served as experimenters.

[3] The skin conductance data will be reported in a separate paper. Here we may simply note that the electrodes were attached to the soles of the feet rather than elsewhere, to permit free movement of the hands. Long leads permitted the subject to move the necessary distances.

4. *Preparing and Serving the Tea*

Every effort was made to serve the tea in a pleasant fashion. Before the session, a tray was prepared in an adjoining room. In addition to cups, saucers, spoons, sugar, and tea bags, the tray held a glass pot and hot plate. A remote control switch was located close to other controls so that one experimenter could unobtrusively turn on the hot plate 10 minutes before the pause. This was done every day, so the subjects would have no indication whether or not tea was to be served. This precaution, together with the *ABBA* order, was successful in creating the impression that tea was served at the whim of the project director, and that even the experimenters couldn't predict his spontaneous act of generosity.

At the start of the pause, one experimenter brought in the tray. Tea was prepared by pouring boiling water into a cup containing a Lipton tea bag. The strength of the tea and the use of sugar was left up to the subject. None insisted on cream. A second cup was offered, but rarely accepted. During the pause while serving and drinking tea, and during the pause when no tea was served,[4] both experimenters and both subjects engaged in casual conversation. The pause was 15 minutes in duration, allowing time for experimenter or subject to go to the toilet if desired.

5. *Description of Tests and Tasks*

a. Hand steadiness. This test has been widely used in studies of fatigue and drug effects. Both fatigue and caffeine tend to decrease steadiness. The precise physiological significance of steadiness is undoubtedly quite complicated. It is probably most directly related to uneven excitation of the muscles of the hand, and in this sense may be regarded as one index of "jumpiness" (or "jitters"). However, extreme relaxation may also impair control of the hand, through inadequate tonus.

---

[4] The alternative of serving hot water during the control pauses was deliberately rejected; we were measuring the effects of the act of drinking tea in a social setting, and not attempting to analyze the components of this act.

The test used consisted of the subject's holding a metal stylus in the hole of a stainless steel plate set at an angle of 45°. There were four 15-second tests within each test period, two with each hand. Contacts were indicated by the flashing of a light, and duration of contacts was measured in hundredths of seconds with a Springfield timer. A metal nut placed at the end of the stylus prevented any possibility of the subject's "cheating" by withdrawing the stylus from the hole. Such withdrawals were noted during our preliminary study.

For statistical treatment, the durations of contacts were transformed to square roots, the score for each test period for each subject being the mean of the two 15-minute tests. Data for the two hands were analyzed separately.

b. Simple reaction time to an auditory stimulus. This test measures speed of reaction in an extremely simple situation. As such, it represents an index of alertness. There were 20 presentations of a click and hum (60-cycle A.C. presented through earphones) to which the subject had to respond by snapping a toggle switch. The experimenter said, "Ready," prior to each trial. The foreperiod varied from one to four seconds.

The score for each period for each subject was the mean reaction time for the 20 trials, in hundredths of seconds. Variability (*SD*'s) of the reaction times was also analyzed.

c. Complex reaction time to visual stimuli. This test differs from the preceding one in that it taps a somewhat higher function, i.e., the ability to make rapid discriminations. It was conducted in the same manner as the Simple Reaction Time test, except that the subject was instructed to keep her left hand on the left toggle switch, her right hand on the right toggle switch, and to respond with her left hand to an amber light, and with her right hand to a blue light. Treatment of data was the same as that for simple reaction time.

d. Strength of grip. The subject squeezed the Smedley hand dynamometer four times during each test period, twice with each hand. Data for the two hands were analyzed separately, using the mean of the two squeezes with each hand as the basic score. The units were kilograms.

e. Attention. The test used to measure attention was one requiring sustained concentration on patterns of auditory stimuli. It was based on Test 11, devised by Wittenborn (4). During each test period the subject listened to 120 sets of three digits, presented via tape recording and earphones.[5] For the first two experimental days, rate of digit presentation within a set of three was 92 per minute. One beat separated each set of three digits. For the third and fourth experimental days, rate of digit presentation was 120 per minute. The subject's task was to mark a plus whenever certain combinations of numbers appeared. The score was 120 (maximum possible) minus two times the number of errors made, to correct for guessing.

f. Subjective ratings of fatigue and tension. During each test period the subject was asked: "On a scale from one to five, where *one* represents little or no fatigue and *five* represents extreme fatigue, how fatigued or tired do you feel?" A comparable question was asked concerning tension during each test period.

g. Long division. The three 15-minute periods of long division were inserted in the session primarily to induce fatigue rather than to serve as a test. The dividends of the examples were six-digit numbers; the divisors, three-digit numbers. Examples were never repeated.

C. RESULTS

Table 3 summarizes statistical analyses of the findings. Note that the scores in this table include only measures taken during the last two experimental periods, where practice effects appeared absent. Note further that the scores in the body of the table represent net gains or losses as a result of pause-with-tea. That is, each individual's pre-pause test score (Test Period 2) was taken as a base for computing her post-pause gains (or losses). The entries in the table represent the differences in such gains (or losses) between tea and no-tea days. The first entry in the table, labeled "Reference Score," gives *mean* pre-pause score (tea and no-tea days combined). This mean value was not used in the computation, but is presented merely as an aid in evaluating the magnitude of the net gains (or losses).

1. *Strength of Grip*

Our preliminary study on six subjects failed to show that tea had a significant effect on strength of grip in either hand. This test was retained in the final battery, since it seemed desirable to have some measure of gross physical effort. Table 3 shows clearly that tea had no significant effect on strength of grip.

---

[5] We wish to thank Dr. J. R. Wittenborn for making available his original recording as a model.

TABLE 3
DIFFERENCES BETWEEN GAINS OR LOSSES IN TEA VS. NO-TEA DAYS

Test	Reference score*	Immediate tests M-diff.**	Immediate tests *p*	First delayed tests M-diff.**	First delayed tests *p*	Second delayed tests M-diff.**	Second delayed tests *p*
Simple Reaction Time							
Speed	30.4	3.90	<.01	3.45	<.01	3.79	<.01
Variability (*SD*)	5.58	—0.39	>.05	—0.05	>.05	0.55	>.05
Complex Reaction Time							
Speed	36.9	1.84	<.01	1.67	.05-.02	2.46	<.01
Variability (*SD*)	5.57	—0.69	>.05	—0.49	>.05	0.58	>.05
Steadiness							
Preferred Hand	22.6	—1.04	>.05	—2.47	>.05	—0.76	>.05
Non-preferred	24.0	—0.20	>.05	—0.19	>.05	—2.22	>.05
Grip							
Preferred Hand	25.8	0.07	>.05	0.68	>.05	0.60	>.05
Non-preferred	21.7	0.04	>.05	0.31	>.05	0.42	>.05
Attention	104.8	—1.27	>.05	1.95	>.05	—0.80	>.05
Long Division							
Amount	39.7	—0.50	>.05	1.65	>.05	—	—
Number Correct	32.3	—0.09	>.05	—0.55	>.05	—	—
Fatigue (Subjective)	1.68	0.32	>.05	0.41	>.05	0.45	>.05

*Units of measurement for each of the tests were as follows: simple and complex reaction time, hundredths of seconds; steadiness, square root of contact time in hundredths of seconds; grip, kilograms; attention, number right minus two times the number wrong; long division, number of digits in quotient; fatigue, rating scale ranged from 1-5.

**Scores were set up so that a positive difference always indicates a favorable effect of tea.

2. *Hand Steadiness*

In our preliminary series of tests, steadiness of the preferred hand appeared considerably improved after pause-with-tea, showing up most strongly in the delayed test. This was particularly surprising in that previous work indicated a decrease in steadiness as the caffeine in tea presumably took effect. However, the results of the main series do not confirm these preliminary results. As is apparent in Table 3, the mean differences are small and not statistically significant. But both sets of results are in agreement in showing *no loss* of hand steadiness as a result of drinking tea.

3. *Attention*

In the preliminary experiment, this test indicated that tea had a slightly detrimental effect on sustained attention to patterns of numbers presented by means of a tape recorder. The significance of this finding was not clear, particularly because the test was too easy. In the main experiment this test was speeded up to make it more discriminating. Under these changed conditions, the apparent deleterious effect of tea dropped out, the differences being small and not statistically significant (Table 3). It may be noted in passing that any standard test of attention is particularly difficult to administer repeatedly to the same subject. With practice, the subject's adjustments tend to become automatized; the "attention" test is no longer a test of attention.

4. *Simple Reaction Time to Auditory Stimuli*

As in the preliminary tests, simple reaction time showed a consistently favorable effect of tea (Table 3). The net gains from tea were significant beyond the one per cent level of confidence for the immediate, first delayed, and second delayed tests. There was no evidence of changes in variability as a result of tea. Since speed of reaction is widely used as a measure of *alertness,* this favorable effect is of importance.

A breakdown of the simple reaction time data was carried out to determine whether factors other than pause-with-tea were responsible for the over-all effect. Statistical analyses were performed on each of the following categories of subjects: habitual mealtime drinkers, habitual afternoon drinkers, sugar-users, non-sugar users, tea on Day 3 subjects, and tea on Day 4 subjects. All differences were significant beyond the one per cent level of confidence on the immediate post-tea test, despite the reduction in degrees of freedom, but some differences did not maintain this high a level of significance on delayed tests.

The absolute magnitude and direction of the changes for all subjects can be seen in Figure 1. Note that there was a consistent tendency for reaction time to increase from the first to the fifth tests on the no-tea day. This loss in alertness was apparently restored by a pause-with-tea, and the subjects maintained their position during all three post-tea tests. The absolute changes were not great, about 10 per cent in order of magnitude. But it must be remembered that reaction time is a highly stable measure once the major initial practice effect has been overcome. For a correct statement of the percentage differences, we would have to relate the differences to what Piéron (2) has called the *reducible margin* in reaction time, which of course would make the percentage differences much larger.

The graph suggests that there was some mismatch of tea and no-tea scores on pre-pause tests. This difference may represent a slight loss in efficiency on the fourth day; it will be recalled (Table 1) that 14 of the 22 subjects had tea on this day. Both the figure and the analysis of data show the favorable effect on pause-with-tea despite this imbalance.

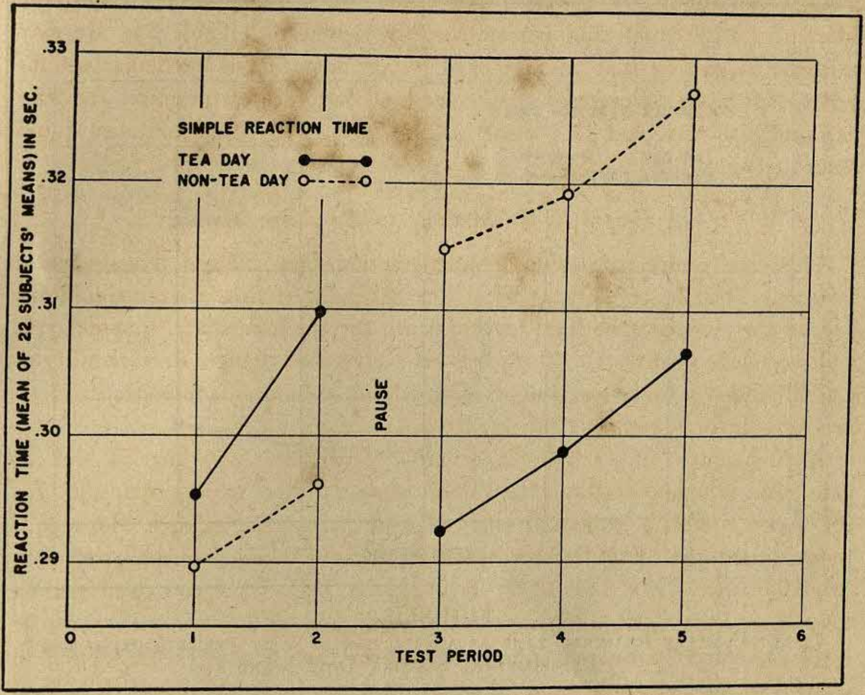


FIGURE 1
MEAN SIMPLE REACTION TIME AS A FUNCTION OF TEST PERIOD DURING THE EXPERIMENTAL SESSIONS, DAYS 3 AND 4

5. *Complex Reaction Time to Visual Stimuli*

It will be seen in Table 3 and Figure 2 that the differences obtained were smaller than those obtained for simple reaction time. However, the favorable effect of tea was significant beyond the one per cent level on the immediate and second delayed tests, and between the two and the five per cent levels on the first delayed test. Breakdown by subject categories showed the same picture found for simple reaction time, except that the differences did not reach as high a degree of statistical significance, particularly on the first delayed test.

Although the results of both preliminary and main experiments leave no doubt that tea had a favorable effect on complex reaction time, there remains the problem as to why the effect was less pronounced than it was for simple reaction time. One possibility is that an increase in level of alertness and the resulting tendency to make a quick response occasionally

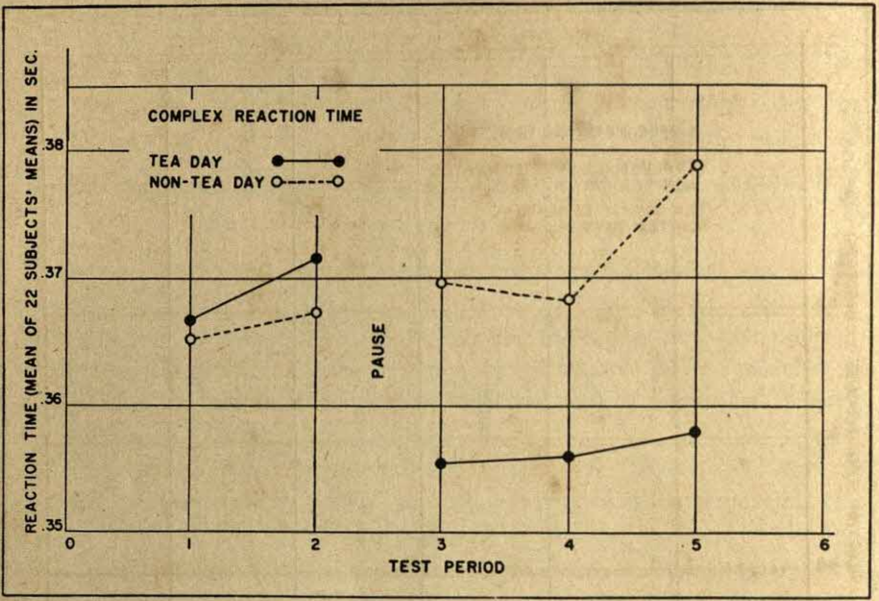


FIGURE 2
Mean Complex Reaction Time as a Function of Test Period During the Experimental Sessions, Days 3 and 4

leads to an actual or threatened error. Such an error would break up the response set, and would be followed by several long reactions. Unfortunately, the number of overt errors was too small to permit a test of this hypothesis. Further work is in progress on this question.

Finally, it should be noted that there is almost no disparity between tea and no-tea days on the pre-pause tests, as there had been for simple reaction time. Since the imbalance in schedule applied equally to both tests, there is the suggestion that level of interest was higher in complex reaction time, thus preventing any decrement due to increasing fatigue or boredom during the series of sessions.

6. *Subjective Ratings of Fatigue and Tension*

The mean gains from pause-with-tea in subjective estimates of fatigue were small, and of doubtful significance from a statistical standpoint. Nonetheless, the plot of mean self-ratings (Figure 3) is highly suggestive in that there was a steady increase in the reported fatigue during the no-tea day, with no break as a result of pause alone. On the other hand, the reports

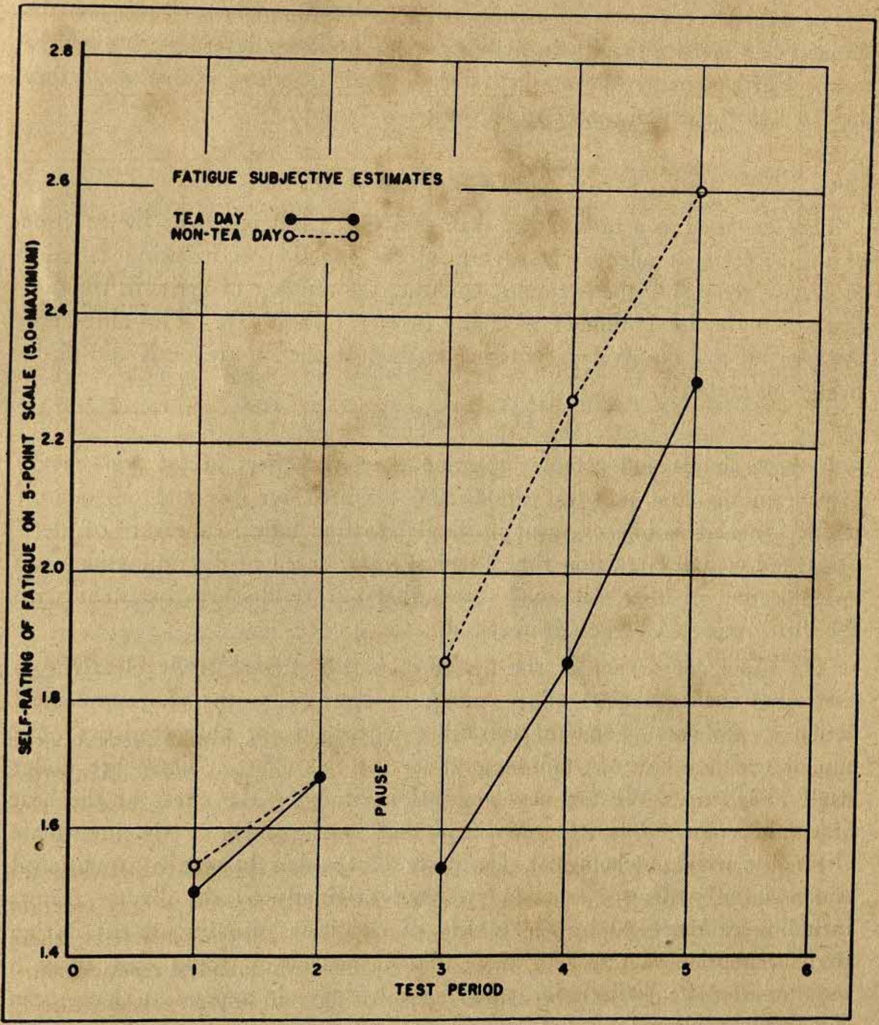


FIGURE 3
Mean of Subjects' Ratings of Fatigue as a Function of Test Period During the Experimental Sessions, Days 3 and 4

of fatigue dropped after pause-with-tea, and stayed below the no-tea day thereafter. The chart exaggerates these differences; the whole change from Test 1 to Test 5 was only one division on a five-point scale. It might have been better to have used a nine-point scale for such judgments, thus increasing the possible precision of the results.

The subjective reports of *tension* were quite unsatisfactory because the subjects had some difficulty in arriving at a consistent interpretation of the term. Further, many of them hesitated to admit tenseness, even though they appeared to the experimenter to be obviously "jumpy."

7. *Long Division*

This was used as a task rather than as a test. Judging from the reaction of the subjects, it adequately performed the function of inducing fatigue.[6] However, we scored the examples, counting the number of digits in the quotients and the total number of digits correct (Table 3). The differences between tea and no-tea days by either method of scoring are small and statistically insignificant.

D. DISCUSSION

In both simple and complex reaction time, the effect of tea was clearly present in the first post-tea period[7] and persisted for one and one-quarter hours. Insofar as improvement in simple reaction time (a measure of alertness) and complex reaction time (a measure of speed of discrimination) are an indication of "lift," it would seem that our results are consistent with the "lift" reported by tea drinkers.

The rapid appearance of the tea effect in the present study clearly suggests that the immediate effect cannot be ascribed to the pharmacological action of caffeine. Some of the other ingredients or characteristics of a cup of tea may be important determiners of this effect. These are under study elsewhere. We are now engaged in studying the effect of the heat of the beverage. But it seems likely that some portion of the immediate effect is purely psychological, i.e., taste of tea has become a conditioned stimulus for evoking a reaction produced originally by the diverse factors including caffeine and by the variety of conditions usually present during the consumption of tea. Although this learned effect would seem to be a most appropriate subject for psychological study, it appears to have been completely neglected.

Cheney (1) did find a hint that the favorable effect of coffee (but not caffeine) on reaction time showed up in 30 minutes, but the improvement was small and was not emphasized. The present study shows that tea has a clear-cut and statistically significant effect on the immediate and other post-pause tests of reaction time.

---

[6] There is no doubt that the session as a whole produced a steady increase in fatigue. This is especially clear in simple reaction time, no-tea day, as shown in Figure 1.

[7] Half of the *S*s completed both Simple and Complex Reaction Time tests within 15 minutes of the first sip of tea; their running mates finished seven minutes later.

E. SUMMARY AND CONCLUSIONS

Twenty-two tea-drinking subjects were run through a five-session cycle of tests to determine some of the immediate and delayed effects of drinking tea in a social situation. During the four experimental sessions, there were two sessions during which the subjects received a pause-without-tea as a control for the two sessions during which the subjects received a pause-with-tea.

1. Simple auditory reaction time was improved by tea, and the effect persisted through three test periods, the last of which was an hour and a quarter after tea. Complex reaction time to visual stimuli showed a similar but smaller effect. These changes in simple and complex reaction time were in agreement with data from a preliminary study on six subjects. Such improvement may be regarded as objective evidence that a cup of tea affords an immediate as well as a delayed "lift" without inducing secondary depressing effects in our subjects, under the conditions studied.
2. Hand steadiness showed no significant gain or loss from tea. This is to be contrasted with the usual report of unsteadiness following caffeine.
3. Tests for attention and strength of grip showed no clear effect after tea. Long division served well as a fatiguing task but its performance was not clearly changed by tea.
4. Subjects self-ratings of fatigue varied as expected, but differences between tea and no-tea days were not statistically significant. Ratings of tension were inconsistent, due to the ambiguity of the term.
5. The increased alertness and speed of discrimination found in the delayed tests of simple and complex reaction time may be attributed to the diverse ingredients in tea, including caffeine. But the presence of a comparable gain on the immediate post-pause tests of these functions is probably partly physiological, due to factors other than caffeine, and partly psychological or learned.

REFERENCES

1. CHENEY, R. H. Reaction time behavior after caffeine and coffee consumption. *J. Exper. Psychol.*, 1936, **19**, 357-369.
2. PIÉRON, H. De la loi du variation des temps de latence en fonction des intensités excitatrices pour les sensations auditive. *C. R. Soc. de Biol.*, 1919, **82**, 1116-1118.

3. Tufts College, Institute for Applied Experimental Psychology, Special Devices Center, Technical Report No. SDC 199-1-2. *Handbook of Human Engineering Data,* 2d ed., 1951.
4. Wittenborn, J. R. Factorial equations for tests of attention. *Psychometrika,* 1943, 8, 19-35.

Psychological Laboratory
Brown University
Providence 12, Rhode Island